

Streptophyte evolution: plastidial signals and core abiotic stress physiology of the closest algal relatives to embryophytes

Dissertation

For the award of the degree
“*Doctor rerum naturalium*” (Dr. rer. nat.)
of the Georg-August-Universität Göttingen

within the doctoral program
“Microbiology and Biochemistry”
of the Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular
Biosciences (GGNB)

Submitted by
Janine Maria Regina Fürst-Jansen
from
Neuss

Göttingen 2023

Thesis Advisory Committee

1st Referee: Prof. Dr. Jan de Vries

Department of Applied Bioinformatics, Institute for Microbiology and Genetics,
University of Göttingen

2nd Referee: Prof. Dr. Ivo Feußner

Department of Plant Biochemistry, Albrecht-von-Haller-Institute for Plant Sciences,
University of Göttingen

Prof. Dr. Jörg Stülke

Department of General Microbiology, Institute for Microbiology and Genetics,
University of Göttingen

Dr. Maike Lorenz

Experimental Phycology and Culture Collection of Algae at the University of Göttingen
(EPSAG), Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen

Members of the Examination Board

1st Referee: Prof. Dr. Jan de Vries

Department of Applied Bioinformatics, Institute for Microbiology and Genetics,
University of Göttingen

2nd Referee: Prof. Dr. Ivo Feußner

Department of Plant Biochemistry, Albrecht-von-Haller-Institute for Plant Sciences,
University of Göttingen

Further Members of the Examination Board

Prof. Dr. Jörg Stülke

Department of General Microbiology, Institute for Microbiology and Genetics,
University of Göttingen

Dr. Maike Lorenz

Experimental Phycology and Culture Collection of Algae at the University of Göttingen
(EPSAG), Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen

Prof. Dr. Stefanie Pöggeler

Department of Genetics of Eukaryotic Microorganisms, Institute for Microbiology and
Genetics, University of Göttingen

Prof. Dr. Rolf Daniel

Department of Genomic and Applied Microbiology, Institute for Microbiology and
Genetics, University of Göttingen

Date of the oral examination: March 23th, 2023

Affidavit

I hereby declare that this dissertation has been written independently and with no other sources and aids than quoted. This dissertation has not been submitted in the same or similar form to other institutions. I have not previously failed a doctoral examination procedure.

Göttingen. 24. November 2023

Janine Maria Regina Fürst-Jansen

Table of Content

LIST OF IMPORTANT ABBREVIATIONS

SUMMARY

1	INTRODUCTION	1
1.1	TWO MAJOR EVENTS IN PLANT HISTORY	1
1.2	STREPTOPHYTE ALGAE: THE CLOSEST ALGAL RELATIVES TO EMBRYOPHYTES	3
1.3	STRESS RESPONSES IN LIGHT OF STREPTOPHYTE EVOLUTION	8
1.3.1	Examples of successful abiotic stress response throughout the green lineage	8
1.4	SPECIALIZED PLANT METABOLISM IN STREPTOPHYTE ALGAE	15
1.4.1	The Phenylpropanoid Pathway: The evolution of a hallmark land plant pathway for specialized metabolites	15
1.4.2	The evolutionary conservation of lipid droplet formation	18
2	AIM OF THIS THESIS	21
3	CHAPTER I: A BROAD VIEW OF ABIOTIC STRESS RESPONSE IN STREPTOPHYTE ALGAE	23
3.1	PUBLICATION I: EVO-PHYSIO: ON STRESS RESPONSES AND THE EARLIEST LAND PLANTS	23
3.2	PUBLICATION II: SUBMERGENCE OF THE FILAMENTOUS ZYGNEMATOPHYCEAE <i>MOUGEOTIA</i> INDUCES DIFFERENTIAL GENE EXPRESSION PATTERNS ASSOCIATED WITH CORE METABOLISM AND PHOTOSYNTHESIS	41
3.3	PUBLICATION III: ENVIRONMENTAL GRADIENTS REVEAL STRESS HUBS PREDATING PLANT TERRESTRIALIZATION	65
4	CHAPTER II: REBUILDING STREPTOPHYTE EVOLUTION WITH COMPARATIVE GENOMICS	99
4.1	PUBLICATION IV: CROSSROADS IN THE EVOLUTION OF PLANT SPECIALIZED METABOLISM	99

4.2	PUBLICATION V: THE EVOLUTION OF THE PHENYLPROPANOID PATHWAY ENTAILED PRONOUNCED RADIATIONS AND DIVERGENCES OF ENZYME FAMILIES	125
4.3	PUBLICATION VI: DIFFERENT PATTERNS OF GENE EVOLUTION UNDERPIN WATER-RELATED INNOVATIONS IN LAND PLANTS	157
4.4	PRE-PRINT VII: CHROMOSOME-LEVEL GENOMES OF MULTICELLULAR ALGAL SISTERS TO LAND PLANTS ILLUMINATE SIGNALING NETWORK EVOLUTION	163
4.5	PUBLICATION VIII: UNEXPECTED CRYPTIC SPECIES AMONG STREPTOPHYTE ALGAE MOST DISTANT TO LAND PLANTS	229
5	DISCUSSION	241
5.1	ON CHAPTER I: A BROAD VIEW OF ABIOTIC STRESS RESPONSE IN STREPTOPHYTE ALGAE	241
5.1.1	Environmental changes reveal fast-responding adjustment of photosynthetic mechanisms in <i>Mougeotia</i> (Publication II)	242
5.1.2	In-depth multifaceted analysis reveals plastid-derived molecular programs in <i>Mesotaenium</i> (Publication III)	245
5.1.3	Concluding remarks on Chapter I	253
5.2	ON CHAPTER II: COMPARATIVE GENOMICS CAN UNCOVER CONSERVED GENETICAL FEATURES FACILITATING TERRESTRIALIZATION	253
5.2.1	The deep evolutionary roots of the phenylpropanoid pathway (Publication V)	255
5.2.2	Chromosome-scale genome reveal co-option networks connected to programs for multicellularity (Preprint VII)	258
5.2.3	Comparative analyses unravel diversity in Chlorokybophyceae (Publication VIII)	261
5.2.4	Concluding remarks on Chapter II	262
6	OUTLOOK	263
7	REFERENCES	267
8	ACKNOWLEDGMENTS	285
9	CURRICULUM VITAE	287

Publications and pre-prints included in this thesis

- I **Fürst-Jansen JMR**, de Vries S, de Vries J. (2020) Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany*. 71: 3254–3269
- II **Fürst-Jansen JMR**, de Vries S, Lorenz M, von Schwartzberg K, Archibald JM, de Vries J. (2021) Submergence of the filamentous Zygnematophyceae *Mougeotia* induces differential gene expression patterns associated with core metabolism and photosynthesis. *Protoplasma*. 259: 1157–1174
- III Dadras A, **Fürst-Jansen JMR**, Darienko T, Krone D, Scholz P, Sun S, Herrfurth C, Rieseberg TP, Irisarri I, Steinkamp R, Hansen M, Buschmann H [8 co-authors] de Vries J (2022) Environmental gradients reveal stress hubs predating plant terrestrialization. *Nature Plants*. 9: 1419-1438: <https://doi.org/10.1038/s41477-023-01491-0>
- IV Rieseberg TP, Dadras A, **Fürst-Jansen JMR**, Dhabalia Ashok A, Darienko T, de Vries S, Irisarri I, de Vries J. (2022) Crossroads in the evolution of plant specialized metabolism. *Seminars in Cell and Developmental Biology*. 134: 37-58
- V de Vries S, **Fürst-Jansen JMR**, Irisarri I, Ashok AD, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J. (2021) The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *The Plant Journal*. 107: 975-1002
- VI **Fürst-Jansen JMR**, de Vries S, Irisarri I. (2022) Different patterns of gene evolution underpin water-related innovations in land plants. *New Phytologist*. 235: 380–383
- VII Feng X, Zheng J, Irisarri I, Yu H, Zheng B, Ali Z, de Vries S, Keller J, **Fürst-Jansen JMR**, Dadras A [38 co-authors] de Vries J, Yin Y (2023) Chromosome-level genomes of multicellular algal sisters to land plants illuminate signaling network evolution. Prepared for submission, Biorxiv: <https://doi.org/10.1101/2023.01.31.526407>
- VIII Irisarri I, Darienko T, Pröschold T, **Fürst-Jansen JMR**, Jamy M, de Vries J. (2021) Unexpected cryptic species among streptophyte algae most distant to land plants. *Proceedings of the Royal Society B: Biological Sciences*. 288: 20212168

In this thesis the publications and pre-prints listed above will be referred to by “**Publication/pre-print I, II, III, IV, V, VI, VII, VIII**”.

List of important abbreviations

4CL	4-COUMARATE-COA LIGASE
ABA	Abscisic acid
BSC	Biological Soil Crusts
C4H	CINNAMATE 4-HYDROXYLASE
CAD	CINNAMYOL ALCOHOL DEHYDROGENASE
CAS	CYCLOARTENOL SYNTHASE
CCR	CINNAMOYL-COA REDUCTASE
CLPPs	CASEINOLYTIC PROTEASES
CLPRs	CASEINOLYTIC PROTEASES-RELATED
COP1	CONSTITUTIVELY PHOTOMORPHOGENIC1
CoRR	Co-Location of Redox Regulation
EGT	Endosymbiotic Gene Transfer
ELIP	EARLY LIGHT INDUCED PROTEINS
FtsH	FILAMENTOUS TEMPERATURE SENSITIVE HAGE
GLK1	GOLDEN-LIKE1
GUN2	GENOMES UNCOUPLED2
HGT	Horizontal Gene Transfer
HPLC	High Performance Liquid Chromatography
HSD	HYDROXYSTEROID DEHYDROGENASES
HSF	HEAT-STRESS RESPONSIVE TRANSCRIPTION FACTORS
HSP	HEAT SHOCK PROTEIN
HY5	ELONGATED HYPOCOTYL 5
HY5	HY5: ELONGATED HYPOCOTYL5
ITS	Internal transcribed spacer
KEGG	Database (Kyoto Encyclopedia of Genes and Genomes)
LC-MS	Liquid chromatography-mass spectroscopy
LD	Lipid droplets
LDAP	LD-ASSOCIATED PROTEIN
LHCSR	Light Harvesting-Like Protein Stress Related
LHCs	Light-Harvesting Complexes
LysM	Lysin Motif
MAA	Mycosporine-like Amino Acids
NPQ	Non-photochemical quenching
OLE 7	OLEOSIN
PAL	PHENYLALANINE-AMMONIA LYASE
PAM	Pulse-Amplitude Modulation
PAR	Photosynthetically Active Radiation
PCA	Principal Component Analysis
PP2C	PHOSPHATASE 2C
PSBS	Photosystem II Subunit S
PYR	PYRABACTIN RESISTANCE 1
RCAR	REGULATORY COMPONENTS OF ABA RECEPTOR
RL	Receptor-Like Kinases
ROS	Reactive Oxygen Species
SLAC	S-Type Anion Channels
SSU	Small Subunit
SnRK	SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASE
TAG	Triacylglycerol
TF	Transcription Factor

UVR	Ultraviolet Radiation
UVR8	UV RESISTANCE LOCUS 8
WGCNA	Weighted Gene Co-Expression Network Analysis
bZIP	BASIC LEUCINE ZIPPER
dn-OPDA	DINOR-12-OXO-10,15(Z)-PHYTODIENOIC ACID
pTAC6	PLASTID TRANSCRIPTIONALLY ACTIVE 6
qE	Energy-dependent Quenching
qT	State-transition Quenching
qZ	Zeaxanthin-dependent Quenching

Summary

Plant terrestrialization facilitated the emergence of the enormous and diverse macroflora. It changed the face of our planet. On this singular evolutionary event, which occurred approximately 550 million years ago, land plants emerged from within the clade of Streptophyta. Many facets of this event remain obscure and are among the central questions of plant evolutionary research. One of the all-time general interest questions is: why did plant terrestrialization happen only once in the evolution of the green lineage? One attempt to answer this question is to take a closer look at the factors that might have enabled this event.

The earliest land plants likely had to overcome a barrage of environmental challenges to thrive in the terrestrial habitat. To do so, they had to possess the molecular toolkit for responding to terrestrial stressors effectively. Understanding facets of this molecular core-toolkit was therefore the aim of this thesis. I compared a set of molecular traits shared by land plants and their closest algal relatives, the streptophyte algae, to infer the trait repertoire that the last common ancestor of land plants and streptophyte algae possessed. This inference can give an idea of the molecular trait setup of the earliest land plants. In this thesis, the focus was on stress-responsive traits throughout the paraphylum of streptophyte algae, with a particular focus on the class of Zygnematophyceae—the phylogenetically closest streptophyte algal relatives to land plants. The algal progenitors of land plants were likely pre-adapted to terrestrial environments. Indeed, a mosaic of core-stress response mechanisms that are well studied in land plants can be found in the streptophyte algal lineage as well and some of them were studied in detail in this thesis using a comparative approach.

A successful response to stress is crucial for plant survival and involves sensing, fine-tuned signaling and responding by the production of suited compounds. In the plant cell all these processes are connected to the plastid. This thesis showed that plastid-derived signals are at the heart of streptophyte algal stress response as well—and modulate algal physiology under environmental challenges. Three representatives of the Zygnematophyceae stood in the focus of this thesis. Underlying molecular co-expression networks intertwined with plastid-signaling cascades stood out in two species (*Mesotaenium* and *Zygnema*) and were hinted at in the filamentous Zygnematophyceae *Mougeotia*. Thus, these conserved plant molecular programs that these algae share with land plants might have also played a role during plant terrestrialization.

The comparative approach is at the heart of this thesis. In the context of plant terrestrialization-related questions the application of comparative genomics can be extremely useful. One characteristic of land plant stress response that might have aided the earliest land plants in their conquest is the production of a diverse set of specialized metabolites. Here, the plastid is also deeply involved as pathways involved in special metabolite production, like the shikimate or salicylic acid pathway, are plastid-located and precursors for specialized cytosolic pathways stem from the plastid. Comparative phylogenomics performed as part of this thesis shed light on the evolution of one of those pathways producing specialized metabolites, the phenylpropanoid pathway. The underlying chassis of enzymes of this hallmark land pathway brings forth a diverse set of specialized compounds and is shaped by environmental conditions. The enzyme families that make up this pathway can be highly promiscuous and experienced pronounced radiations through the evolution of the green lineage. Intriguingly, sister to these gene families are enzyme candidates in streptophyte algae, hinting at a possible involvement during plant terrestrialization. Lastly, this thesis showed that comparative genomics can also contribute to solving streptophyte algae diversity, a topic that is — for some classes of this grade — still barely investigated. However, unraveling streptophyte algal diversity is important for macroevolutionary studies that aim to understand characteristics of the streptophyte ancestor.

Altogether, in this thesis a broad molecular evolutionary approach is pursued. This approach is a combination of stress response analyses and phylogenomics with the goal to shed light on conserved mechanisms that facilitated plant terrestrialization.

1 Introduction

1.1 Two major events in plant history

Photosynthesis shaped the face of the earth as we know it today. This essential mechanism in plants is only possible because of the endosymbiotic engulfment of a cyanobacterium by a heterotrophic eukaryote at least 1.2 billion years ago (Parfrey *et al.*, 2011). There are still many open questions regarding this event. For starters, it is still debated whether this cyanobacterial progenitor of the first plastids was a fresh-water or marine living organism (de Vries and Archibald, 2018, Ponce-Toledo 2017, Helmchen *et al.*, 1995). While many aspects of this event remain unknown, what we know is that this primary endosymbiosis event led to the first photosynthetically active eukaryotes: the Archeplastida (Archibald 2015, Martin *et al.* 2015, de Vries and Gould, 2018). During the evolution of the Archeplastida the genes from the cyanobacterial endosymbiont were mostly lost while the nuclear genome of the host became more complex. This transfer of genes is dubbed endosymbiotic gene transfer (EGT) (Martin and Herrmann 1998, Timmis *et al.*, 2004). Only a few relevant genes were retained in the plastid genome. This can be explained by the CoRR-hypothesis (co-location of redox regulation) which describes a redox-reaction based gene regulation which allows the optimal adaption to changing environmental conditions (Allen *et al.*, 2015).

The monophyletic group of the Archeplastida eventually split again into three major lineages: the Chloroplastida, the Rhodophyta and the Glaucophyta (Roudríguez-Ezpeleta *et al.*, 2005). The emergence of the Chloroplastida set the stage for the second major event in plant history: the global conquest of land by the green lineage, also dubbed as plant terrestrialization. Roughly a billion years ago the Chloroplastida split into the Chlorophyta and the Streptophyta (Morris *et al.*, 2018). Only from a single clade within the streptophyte lineage, the Embryophyta that will be henceforth called land plants in this thesis, eventually emerged in a singular event that happened approximately 550 million years ago (Morris *et al.*, 2018, Harris *et al.*, 2022); this led to the enormous and diverse macroflora we know today and facilitated the present-day atmospheric oxygen levels (Lenton *et al.*, 2016).

Certain characteristics made the ancestor of the earliest land plant already pre-adapted to the terrestrial environment. These characteristics are of morphological, physiological, and biochemical nature (Becker and Marin 2009). Additionally, the environment in which this ancestor dwelled; moist habitats that frequently encountered wet to dry periods fostered an environment for the emergence of adaptations to the terrestrial habitat.

Fresh water habitats as opposed to marine habitats are furthermore considered to be an advantageous factor for water-to-land transition because they present a gentler ecological gradient (Becker and Marin, 2009, Harholt *et al.*, 2016). The biotic environment aided the ancestor of land plants in its conquest as well. The formation of symbiotic relationships with soil bacteria, first and foremost the formation of arbuscular mycorrhiza, is a major factor for water to land transition (Delaux *et al.*, 2013). Furthermore, the symbiotic exchange between soil dwelling bacteria with the first hypothetical lands was also coupled with a horizontal gene transfer (HGT). Genes that the plant host gained are major players in land plant metabolic pathways and further supported the survival of the earliest land plant in the terrestrial environment (see Gao, 2021 for an overview, Cheng *et al.*, 2019).

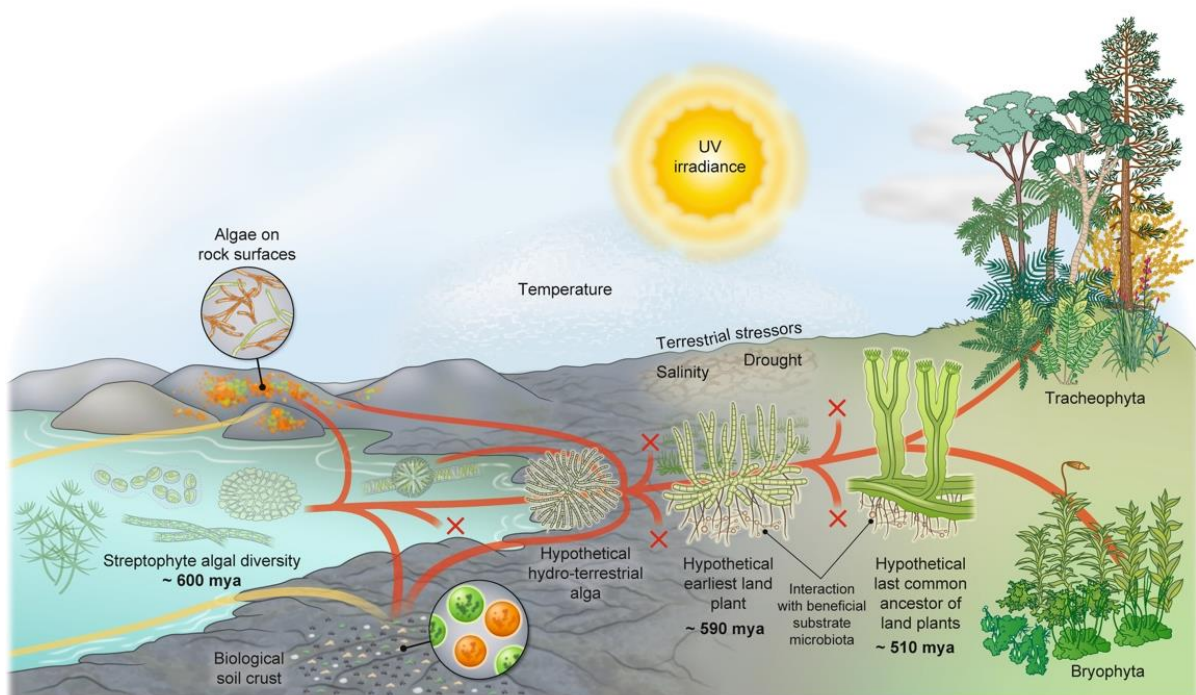


Figure 1: From water to land: A hypothetical reconstruction of plant terrestrialization. A schematic on how the earliest land plant might have looked like. Starting with a diverse set of streptophyte algae approx. 600 million years ago, living in fresh water, on rock surfaces and forming biological soil crusts (BSCs) (yellow lines). Red lines represent the evolutionary trajectory from freshwater dwelling algae to land-dwelling embryophytes. 'X' labels represent lineages that have become extinct throughout evolution. Eventually, the first hydro-terrestrial alga conquered land and the earliest land plant emerged approx. 590 million years ago. This land plant likely lived in an everchanging environment and faced terrestrial stressors such as UV, irradiance, and frequent drought periods. The beneficial interaction with substrate bacteria is also hypothesized. Approx. 510 million years the hypothetical last common ancestor of land plants emerged from which the Bryophyta and Tracheophyta evolved (**Publication I**).

Even if the factors mentioned above are considered to have played major roles during plant terrestrialization there are still many unsolved mysteries surrounding this singular event. To solve some of those mysteries and shed light which traits enabled the success of land plants one must turn to their closest algal relatives: the streptophyte algae. The following section will give an insight on the characteristics of these land plant relatives.

1.2 Streptophyte Algae: the closest algal relatives to embryophytes

The Viridiplantae are subdivided into two major divisions: Chlorophyta and Streptophyta. The Chlorophyta includes prominent core chlorophytes like the model organism *Chlamydomonas reinhardtii*. The Streptophyta are the sister clade to Chlorophyta and encompass the streptophyte algae and the monophyletic Embryophyta. The paraphyletic streptophyte algae comprise six classes: Mesostigmatophyceae, Chlorokybophyceae, Klebsormidiophyceae, Charophyceae, Coleochaetophyceae, and the next closest algal relatives to the Embryophyta: The Zygnematophyceae (Figure 2). Streptophyte algae were already pre-adapted to terrestrial conditions (Becker and Marin, 2009). Living in diverse environments like fresh water as well as terrestrial habitats is hereby only one intriguing feature (Lewis and McCourt, 2004). The following paragraph will give a brief insight into the characteristics of the six currently recognized classes of streptophyte algae and examples of stress-responsive properties that help them fight against environmental stressors.

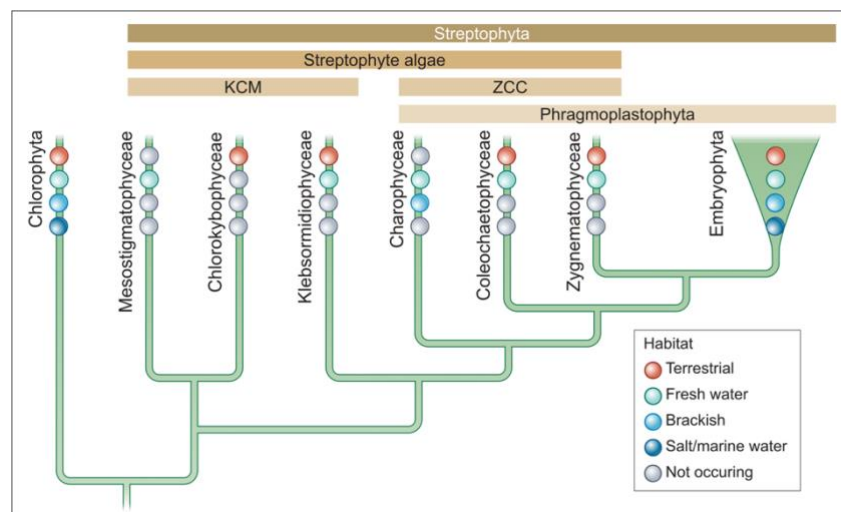


Figure 2: Phylogenetic relationships of Streptophyta and their habitats. The clade Chloroplastida consists of the core green algae (Chlorophyta) and the monophylum Streptophyta which encompasses the monophyletic Embryophyta (land plants) and the paraphyletic streptophyte algae. Streptophyte algae consist of six different classes which can be divided into two different grades: the basal-branching Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae compose the KCM grade. The Charophyceae, Coleochaetophyceae, and the Zygnematophyceae make up the ZCC grade and form the phragmoplast bearing Phragmoplastophyta in a monophylum together with the Embryophyta. Different habitats are shown as dots with corresponding colors (modified from **Publication I**, Chlorokybophyceae habitat is updated according to **Publication VIII**).

At the current state of knowledge, the class Mesostigmatophyceae encompasses two species. The most prominent is *Mesostigma viride*, a fresh water dwelling unicellular flagellate (Marin and Melkonian, 1999), for which the genome was sequenced in 2020 (Wang *et al.*, 2020) is one of them. Interestingly, these recent genome analyses suggest that the ancestor of *M. viride* lost molecular traits that have been associated with subaerial/terrestrial environments, possibly due to its fresh water habitat (Wang *et al.*, 2020).

Sister group to Mesostigmatophyceae are the Chlorokybophyceae. This rare basal line of streptophyte algae is distributed exclusively in terrestrial habitats and originally included only one species: *Chlorokybus atmophyticus*. An in-depth phylogenomic analysis of all available strains in public culture collections world-wide demonstrated that there are at least five species (*C. atmophyticus*, *C. melkonianii*, *C. bremeri*, *C. riethii*, *C. cerfii*) forming a cryptic species complex with the genus *Chlorokybus* (**Publication VIII**). Opposite to *Mesostigma*, *Chlorokybus* inhabits subaerial habitats like soil or rock surfaces (Geitler 1942). *Chlorokybus* forms sarcinoid cell packages, with each cell having its own cell wall (**Publication VIII**, Wang *et al.*, 2020).

Klebsormidiophyceae form the KCM grade with the two priorly discussed streptophyte algal classes. The class of Klebsormidiophyceae is widely distributed and comprises at least 5 different genera: *Klebsormidium*, *Interfilum*, *Entransia*, *Hormidiella* and *Streptosarcina* (Mikahailyuk *et al.*, 2018, Mikahailyuk *et al.*, 2008). Klebsormidiophyceae exhibit unicellular as well as filamentous growth and are found in almost all non-marine habitats across the planet (Mikahailyuk *et al.*, 2015), living in terrestrial as well as fresh-water habitats. In 2014, Hori *et al.* set a milestone by publishing the first genome of a streptophyte alga, the genome of *Klebsormidium nitens* (at the time erroneously called *Klebsormidium flaccidum*; Hori *et al.*, 2014). Studies on various Klebsormidiophyceae species reveal remarkable features that aid them in facing different abiotic stressors. Hartmann *et al.* (2020) found mycosporine-like amino acids (MAAs) in the genera *Interfilum* and *Klebsormidium* that have a photoprotective function under UV radiation (see more details on MAAs in section 1.3.1) (Hartmann *et al.*, 2020). Furthermore, Klebsormidiophyceae show an exceptional desiccation tolerance owing to different attributes like their modified cell walls but also due to the formation of biological soil crusts (BSCs) that protect against water loss and high irradiance (Holzinger and Karsten 2013, Herburger and Holzinger 2015, Holzinger *et al.*, 2011).

The higher branching ZCC grade encompasses the classes Charophyceae, Coleochaetophyceae and the Zygnematophyceae. The latter are the closest algal relatives to land plants based on phylogenetic analyses (Leebens-Mack *et al.*, 2019, Wickett *et al.*, 2014). Together, the ZCC grade and the Embryophyta form the monophylum Phragmoplastophyta (Lecointre and Le Guyader, 2006).

As the name suggests, the algae in this clade encompasses plants that form a phragmoplast, a plant cell-specific structure during late cytokinesis which helps in the formation of a cell plate that later builds up to the cell wall (Buschmann and Zachgo 2016, see also discussions in Nishiyama *et al.*, 2018). Charophyceae (as well as Coleochaetophyceae) were considered for a long time to be the algal ancestors of embryophytes because of their complex reproductive stages (Karol *et al.*, 2001, Pringsheim 1862, van den Hoek *et al.*, 1995). This algal class consists of six genera: *Chara*, *Nitella*, *Tolypella*, *Nitellopsis*, *Lychnothamnus* and *Lamprothamnion* (Krause 1997, Hall and McCourt, 2017). The best-studied species from this class is *Chara braunii*, for which the genome has been available since 2018 and was the second published streptophyte algal genome after *Klebsormidium* (Nishiyama *et al.*, 2018). The genome revealed many land plant-like features that likely aided the algal ancestor in its conquest of land, like a genetic repertoire for retrograde signaling or LysM (lysine motif) receptor-like kinases (RLKs) involved in plant signaling with symbionts or pathogens (Nishiyama *et al.*, 2018).

Table 1: An overview of all sequenced streptophyte algal genomes so far.

Algal class	Species	Strain	Publication
Zygnematophyceae	<i>Mesotaenium endlicherianum</i>	SAG 12.97	Cheng et al., 2019
	<i>Spirogloea muscicola</i>	CCAC 0214	
	<i>Penium margaritaceum</i>	SKD-8 (SAG 2640)	Jiao et al., 2020
	<i>Zygnema circumcarinatum</i>	698-1b, UTEX 1559, UTEX 1560	Feng et al., 2023
	<i>Zygnema cylindricum</i>	698-1a	
	<i>Closterium peracerosum-strigosum-littorale</i> complex	NIES-67	Sekimoto et al., 2022
<i>Closterium peracerosum-strigosum-littorale</i> complex	NIES-68		
Coleochaetophyceae	-	-	-
Charophyceae	<i>Chara braunii</i>	S276	Nishiyama et al., 2018
Klebsormidiophyceae	<i>Klebsormidium nitens</i>	NIES-2285	Hori et al., 2014
Chlorokybophyceae	<i>Chlorokybus atmophyticus</i>	CCAC 0220	Wang et al., 2020
Mesostigmatophyceae	<i>Mesostigma viride</i>	CCAC 1140	

The class Coleochaetophyceae consists of the two genera *Coleochaete* and *Chaetosphaeridium* (Cook *et al.*, 2017). Some algae from the genus *Coleochaete* exhibit a very complex morphology by forming parenchymatous thalli (Graham 1984). These traits however differ along the diversity of *Coleochaete* spp. with the required perpendicular cell division limited to a clade within the genus *Coleochaete* (Delwiche *et al.*, 2002). Coleochaetophyceae are the only streptophyte alga class for which, as of today, no genome is available (Szövényi *et al.*, 2021), even though a genome would help greatly in uncovering underlying molecular traits as *Coleochaete* displays intriguing features like the production of a lignin-like compound (Sørensen *et al.*, 2011, see also section 1.4.1 of this thesis).

The Zygnematophyceae are the algal sister lineage to embryophytes and the most diverse streptophyte alga class with over 4000 described species (Guiry and Guiry, 2021). They comprise of at least 45 genera (Hall and McCourt, 2017).

For a simpler overview Hess *et al.* (2022) recently divided them into a five-order system based on phylogenomic analyses: the Desimidiales, the Spirogyrales, the Zygnematales, the new order Serritaeniales and the Spirogloaeales (Hess *et al.*, 2022). Most of the Zygnematophyceae are unicellular (Hall and McCourt, 2017) which is intriguing because sister to Zygnematophyceae and land plants are the Coleochaetophyceae that show a much more complex, often multicellular morphology; and sister to these are the Charophyceae such as the macroscopic *Chara* (see also Wickett *et al.*, 2014). The assumption that the degree of morphological complexity increases the closer a streptophyte algal class is to land plants is therefore a notion that is too simplistic. Nevertheless, there are also filamentous growing algae belonging to the Zygnematophyceae and interestingly filamentous growth in this class likely emerged at least five times independently (Hess *et al.*, 2022). Zygnematophyceae are also known as conjugating green algae due to their class-specific ability to reproduce sexually via conjugation (Wodniok *et al.*, 2011, Engler 1892). Finally, yet importantly, the *Penium margaritaceum* genome (Jiao *et al.*, 2020) underpinned the findings that *Penium* possesses a complex land plant-like cell wall, making it an interesting species for cell wall studies in streptophyte algae (Domozych *et al.*; 2007; Sørensen *et al.*, 2011).

The focus of this thesis was on three representatives of the Zygnematophyceae: first and foremost, *Mesotaenium endlicherianum* (Strain no.: SAG 12.97), *Mougeotia* (Strain no.: SAG 164.80 and MZCH240) and *Zygnema* (Strain no.: SAG 698-1b, UTEX 1559 and UTEX 1560). *Mougeotia* and *Zygnema* belong to the order of the Zygnematales, while *Mesotaenium endlicherianum* belongs to the novel order of the Serritaeniales (Hess *et al.*, 2022). Starting with *Mesotaenium*, this unicellular alga dwells in freshwater as well as subaerial habitats (Cheng *et al.*, 2019). It has one or two plate-shaped axial plastids with two pyrenoids and a single nucleus in the middle of the cell (Nägeli 1849). Interestingly, for closely related genus *Serritaenia* the formation of an extracellular UV-protective mucilage has been reported (see also section 1.3.1 on UV radiation), a feature that is unusual for microalgae since they tend to produce intracellular sunscreens instead (Busch and Hess, 2021; Hoffmann 1989). In *Mesotaenium* the formation of mucilage upon environmental stress is barely investigated but has been reported before (Fučíková *et al.*, 2008).

Compared to *Mesotaenium* the genome of the filamentous alga *Mougeotia* has not been sequenced yet. *Mougeotia* has unbranched filaments and an axial located plastid which can turn based on light availability (Kadlubowska 1984). Recently, Permann *et al.* (2021) investigated the underlying structural morphology of zygospore formation via conjugation in *Mougeotia* using Raman imaging and found a combination of lipids, aromatic components and carbohydrate material that speaks of algaenan or sporopollenin-like material (Permann *et al.*, 2021).

Both materials are highly resistant with sporopollenin being the material that constitutes the outer walls of spores and pollen (see also Montgomery *et al.*, 2016).

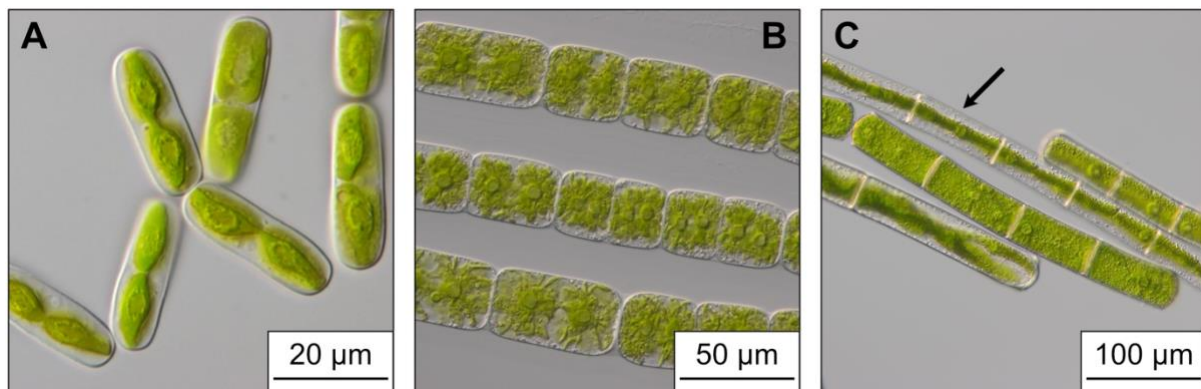


Figure 3: Diverse morphology of three Zygnematophyceae. A) the unicellular alga *Mesotaenium endlicherianum* SAG 12.97, B) the filamentous alga *Zygnema circumcarinatum* SAG 698-1b and C) the filamentous alga *Mougeotia sp.* SAG 11.96, with plastid reorientation shown (arrow). Pictures are taken with an Olympus BX-60 microscope. Scale bars are shown at the bottom.

The filamentous alga *Zygnema* is a very intriguing streptophyte algal genus as it is broadly talented regarding the response to different stresses (see also de Vries *et al.*, 2018; Rippin *et al.*, 2017). Therefore having the genomes of *Zygnema circumcarinatum* and *Zygnema cf. cylindricum* now at hand is a huge advantage (**Publication VII**). The genus *Zygnema* has unbranched filaments, enclosed by a mucilage shield and two stellate chloroplasts per cell with central pyrenoid each. The nucleus lies in between these two plastids, connected by a cytoplasmic bridge (Figure 3, Kadlubowska, 1984). There have been multiple reports on the robustness of *Zygnema* against diverse abiotic stressors. *Zygnema* has a striking tolerance to desiccation as photo-physiological and transcriptomic data on *Zygnema circumcarinatum* exposed to desiccation stress suggest (Rippin *et al.*, 2017). Furthermore, photoprotective UV-absorbing phenolic compounds have been found in Arctic and Antarctic strains of *Zygnema* upon exposure to UV-radiation (Pichrtová *et al.*, 2013; Holzinger *et al.*, 2018). Next to UV-radiation these polar dwelling species are exposed to a combination of stressors with mutual dependencies, such as, freezing, desiccation and osmotic stress. To preserve when facing this cocktail of stressors, they can produce pre-akinetes which contain accumulating lipid bodies and can stay in this stadium for months (Pichrtová *et al.*, 2016; Pichrtová *et al.*, 2013).

This short account of stress-relevant features across the small sub-sample of the biodiversity of streptophyte algae already illustrates specific and environment-adapted mechanisms.

The next section will provide a closer look **a)** against which terrestrial stressors some of these mechanisms are effective against and **b)** how well conserved they are across the green lineage, possibly providing a means to infer their role in plant terrestrialization.

1.3 Stress responses in light of streptophyte Evolution

To reconstruct plant terrestrialization events is a difficult nearly impossible task. Yet, there are studies that aim to infer specific aspects of this fateful event and the factors that played into it. What can be assumed is that the first land plants had to face rapid changes in the environment — of abiotic and biotic nature — that had to be overcome. The diverse embryophytic flora we see today is testament to that. Comparative analyses of the last years revealed that the algal progenitors of land plants featured some of the molecular tools, that are known to be active under specific environmental conditions in embryophytes (**Publications III; V; VII**; Cheng *et al.*, 2019; Nishiyama *et al.*, 2018; Hori *et al.*, 2014). Therefore, we can assume that they have acted upon adverse conditions during plant terrestrialization. The following section will highlight some of the conserved stress response mechanisms that aid algae and land plants in warding off abiotic as well as biotic stressors.

1.3.1 Examples of successful abiotic stress response throughout the green lineage

A successful stress response in the plant involves always (i) sensing, (ii) fine-tuned signaling and (iii) responding by production of suited compounds. The plastid plays a major role in all three processes, but especially the first by acting as an environmental sensor in a process termed retrograde signaling (see also review Chan *et al.*, 2016). Retrograde signaling involves the communication of the plastid with the nucleus which allows nuclear gene expression to be adjusted according to the sensed environmental conditions by the plastid (Koussevitzky *et al.*, 2007; Susek *et al.*, 1993; Pfannschmidt 2003; Kleine *et al.*, 2021; Richter *et al.*, 2022; de Vries *et al.*, 2016). Many genes involved in land-plant retrograde signaling in processes responding to irradiance or cold stress could also be detected in the ZCC-grade of streptophyte algae (**Publication III**; Nishiyama *et al.*, 2018; Zhao *et al.*, 2019; de Vries *et al.*, 2018). The retrograde signaling cascade is involved in various stress responses besides irradiance and cold response. Some of those responses will be discussed in the sections below.

Irradiance

High irradiance (also defined as PAR; Photosynthetically active radiation, with a spectral range from 400 – 700 nm) as well as UV radiation, have a direct effect on the plastid.

As the organelle that senses and processes radiation through the photosynthetic machinery, the plastid holds a vital role for the plant. One of the main targets of high irradiances is Photosystem II (PSII). The PSII core protein D1 is the protein with the highest turn-over rates in the plastid (Weinbaum *et al.*, 1979). A defect D1 protein can contribute to the forming of reactive oxygen species (short: ROS) by transferring photon energy to chlorophylls and generating reactive triplet chlorophyll ($^3\text{Chl}^*$) (see Figure 4, Müller *et al.*, 2001). While a low level of ROS is important in plant signaling and development (e.g., Dunand *et al.*, 2007, Tsukagoshi *et al.*, 2010), high levels of ROS can cause oxidative damage to DNA, lipids and proteins and can therefore lead to the disintegration of the photosynthetic apparatus (Mittler, 2002). Thus, most stress responses are intertwined with responding to an imbalance on the ROS homeostasis.

There are different photoprotective mechanisms to prevent ROS formation. One well-conserved mechanism is the quenching of excess photon energy through heat in a process called non-photochemical quenching (NPQ) (Müller *et al.*, 2001). This process includes different stages that differ in their reaction times. The fasted component is qE (energy-dependent quenching) which is a pH-dependent mechanism and involves the pH-sensing proteins LHCSR (light harvesting-like protein Stress Related) and PSBS (photosystem II subunit S) (Jahns and Holzwarth 2012; Gerotto and Morosinotto, 2013). PSBS manages the reorganization of the light-harvesting antennas of the photosystems to an energy-dissipating state while LhCSR binds pigments and can dissipate energy actively as heat (Betterle *et al.*, 2009; Bonente *et al.*, 2011). It was long thought that LHCSR is only present in chlorophytes and that PSBS is mainly active in embryophytes (Bonente *et al.*, 2011; Li *et al.*, 2000). Surprisingly, PSBS could be detected in the chlorophyte *Chlamydomonas reinhardtii* as well (Correa-Galvis *et al.*, 2016). In the moss *Physcomitrium patens* both proteins are present and active during NPQ responses suggesting a co-existence during streptophyte evolution (Alboresia *et al.*, 2010; Gerotto *et al.*, 2012). In streptophyte algae evidence for a PSBS-dependent NPQ-response could be detected as well: Interestingly, this PSBS-mechanism could only be detected in the ZCC grade whether (only) an LHCSR-dependent mechanism was found in the more basal-branching algae (Gerotto and Morosinotto, 2013).

Another NPQ mechanism involves the reorganization of the light-harvesting complexes (LHCs) and is dubbed qT (state-transition quenching) (Mullineaux and Allen, 1986). qT is a very prominent quenching mechanism in green algae like *Chlamydomonas* where 80% of the light-harvesting complexes of photosystem II (LHCII) can be translocated along the thylakoid membrane upon high irradiance, therefore reducing light harvesting capacity (Delosme *et al.*, 1996). During this process LHCII is reversibly phosphorylated by the chloroplast protein kinase: STT7 in chlorophytes and STN7 in embryophytes (Grieco *et al.*, 2016).

The reorganization of light-harvesting complexes involves also early light induced proteins (ELIPs) which are bound to chlorophyll *a* and *b* in the LHCs and are very conserved in algae and embryophytes (Heddad and Adamska, 2002). ELIPs have a photoprotective function by relocating chlorophylls in the LHCs and even influence their biosynthesis when the photosynthetic machinery faces excess photon energy (Hutin *et al.*, 2003; Tzvetkova-Chevolleau *et al.*, 2007). ELIPs are not only active during elevated irradiances but also in response to diverse stressors and phytohormone signaling like the cascade mounted by abscisic acid (ABA) in land plants as well as in streptophyte algae (Bartels *et al.*, 1992; Zhuo *et al.*, 2013). Interestingly, their corresponding genes could be detected in the transcriptomes under different environmental conditions apart from high irradiance or desiccation in streptophyte algae that are also relevant for this thesis like *Mesotaenium* (see **Publication III**; Cheng *et al.*, 2019), *Mougeotia* (**Publication II**; de Vries *et al.*, 2020) or *Zygnema* (**Publication VII**; de Vries *et al.*, 2018).

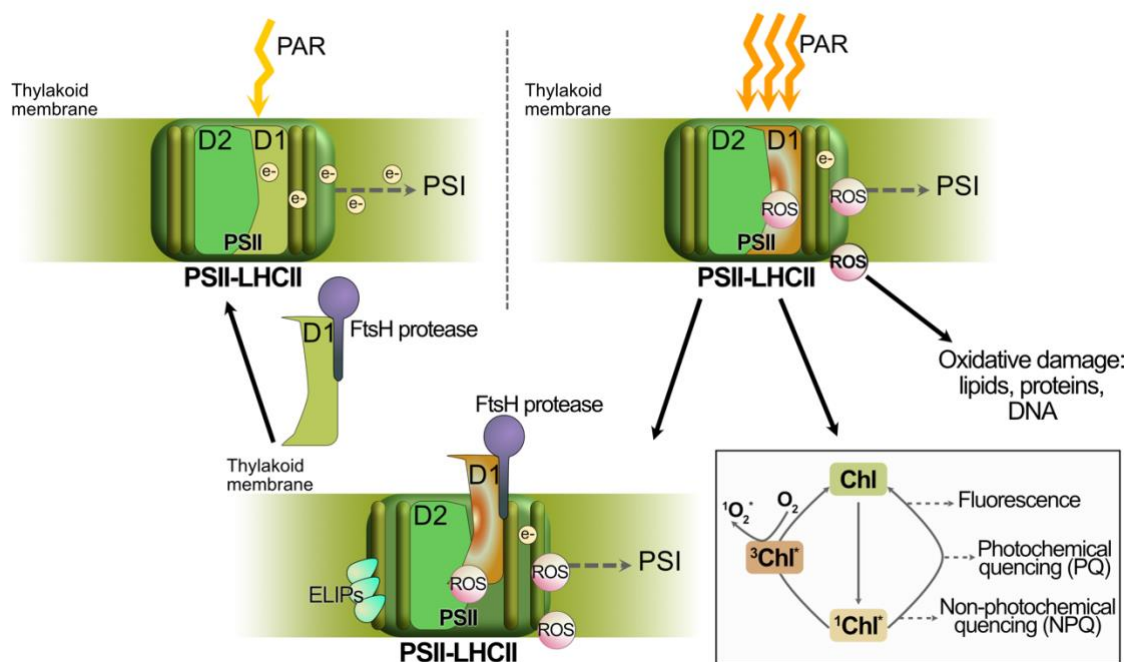


Figure 4: Examples of conserved photoprotective mechanisms in plants. An overview of conserved responses to elevated irradiance (on the top right), which can damage Photosystem II (PSII) core-protein D1 which leads to ROS accumulation. Downstream oxidative damage is then the result. Photoprotective mechanisms that get activated here are for example the NPQ mechanism (on the lower right), the reorganization of light-harvesting complexes of the PSII (LHCII), in which early light-harvesting proteins (ELIPs) are involved, and the removal of damaged D1 protein by the FtsH protease.

Since the PSII core-protein D1 is considerably involved in ROS formation (Kato *et al.*, 2009), plants have a fast-responding mechanism to remove damaged D1 and replace it with new D1. Here the FtsH-protease is involved (filamentous temperature sensitive H), a zinc-metalloprotease that has a quality control function in the plastid by removing damaged D1 and replacing it with a newly synthesized D1 (Bailey *et al.*, 2002; Kato *et al.*, 2009).

Although the evolution of FtsH is complicated (see de Vries *et al.*, 2016), it is intriguing that the plastid-specific *ftsH* gene is plastid-encoded in most chlorophyte as well as in streptophyte algae *Zygnema circumcarinatum* and *Chara vulgaris* while in land plants it is purely nuclear-encoded (de Vries *et al.*, 2013).

The photoprotection of the light-sensitive PSII is also facilitated through Carotenoids. They can directly counteract photooxidative damage by quenching reactive oxygen ($^1\text{O}_2^*$) or triplet Chlorophyll ($^3\text{Chl}^*$) through thermal dissipation of excess light energy (see e.g., Young and Frank, 1996). The deactivation of $^1\text{O}_2^*$ is directly facilitated through β -carotene (Telfer *et al.*, 2002). By-products of ROS-quenching through carotenoids are apocarotenoids, which play a central role in embryophyte stress response by being involved in e.g. phytohormone-mediated- and/or retrograde signaling, also in response to high irradiances (see also **Publication IV**; Zheng *et al.*, 2021). Other carotenoids like xanthophylls are also involved in NPQ-processes and can deactivate triplet chlorophylls. The NPQ-state qZ (zeaxanthin-dependent quenching) involves the reversible conversion of violaxanthin to the excess-energy-quenching zeaxanthin in the violaxanthin cycle (Jahns and Holzwarth, 2012; Niyogi *et al.*, 1997). In streptophyte algae, carotenoid profiles are currently barely investigated. Stamenković and colleagues provide a first insight in pigment characteristics of the Zygnematophyceae *Cosmarium crenatum* under high irradiance conditions (Stamenković *et al.*, 2014). Interestingly, in *C. crenatum* antheraxanthin actively participated in the NPQ-process under high irradiance, displaying an incomplete violaxanthin cycle, an observation that was not seen in other *Cosmarium* strains (Stamenković *et al.*, 2014).

All in all, core-photoprotective measures against high irradiances appear to be rather conserved throughout the streptophyte lineage and involve a fine-tuned molecular adjustment of the photosynthetic machinery.

UV-radiation (UVR)

Next to high irradiance in the visible spectrum solar ultraviolet radiation (UVR), containing mainly ultraviolet A (UV-A, 315 – 400 nm) and ultraviolet B (UV-B, 280-315 nm) is also considered to be a notable terrestrial stressor. Like high PAR intensities, UV radiation (UVR) can greatly impair the photosynthetic apparatus in the plant cell and therefore promotes the production of ROS (Pattison and Davies 2006; Hargreaves *et al.*, 2007). Chlorophyll biogenesis is affected, protein-folding processes are disrupted, and DNA-strand breaks can occur (e.g., Salama *et al.*, 2011; Rastogi *et al.*, 2010).

In the green lineage, the UVR8-mediated (UV RESISTANCE LOCUS8) signaling cascade is a key UV response (Han *et al.*, 2019). In *Arabidopsis thaliana*, the UVR8-COP1 (COP1: CONSTITUTIVELY PHOTOMORPHOGENIC1) signaling pathway counteracts visible light-induced photomorphogenesis in response to low UVB levels (Rizzini *et al.* 2011; Oravecz *et al.*, 2006). In *Arabidopsis*, upon UV-B exposure UVR8 accumulates in the nucleus and interacts with the E3 ubiquitin ligase COP1 (Rizzini *et al.*, 2011). COP1 here acts as a positive regulator for UVB response by promoting the expression of HY5 (HY5: ELONGATED HYPOCOTYL5) which in turn promotes flavonoid accumulation by facilitating the expression of UVB responsive genes (Oravecz *et al.*, 2006; Jenkins *et al.*, 2001).

Another conserved UV-response hypothesized to have played a significant role during plant terrestrialization is the production of UV-absorbing compounds. One famous example for UV-absorption are mycosporine-like amino acids (MAAs). They occur across the green lineage (for an overview see Rozema *et al.*, 2002) and their biosynthesis was inherited by plastid-bearing organisms through their cyanobacterial ancestor via primary endosymbiosis (Whatley, 1981). MAAs can be induced upon UV-B radiation (Xiong *et al.*, 1997), but there are also UV-A-induced MAAs (Sinha *et al.*, 1999). MAAs are passive shielding molecules and can directly absorb UVR and dissipate it as heat (Bandaranayake, 1998). In streptophyte algae, the best characterized MAAs up to date are the MAAs in the Klebsormidiophyceae (Hartmann *et al.*, 2020). Two UV-absorbing MAAs dubbed klebsormidin A and klebsormidin B that absorb at maxima of 324 nm could be successfully isolated from various species of the genera *Interfilum* and *Klebsormidium* and characterized by HPLC (High-performance liquid chromatography) and NMR (Nuclear magnetic resonance) (Hartmann *et al.*, 2020).

UV-absorbing phenolic-compounds are often found in the intracellularly in green algae (Remias *et al.*, 2012a; Remias *et al.*, 2012b). Interestingly terrestrial zygmatophycean species, *Serritaenia*, a sister species to *Mesotaenium* SAG 12.97, can produce extracellular mucilage that contain UV-absorbing phenolics (Busch and Hess, 2021). This mechanism resembles what is known in terrestrial cyanobacteria, which produce extracellular pigments, like the UV-protectant Scytonemin which possesses phenolic subunits (Proteau *et al.*, 1993). A last interesting example of phenolic compound production under UVR exposure was found in different species from *Zygnema* dwelling in Arctic and Antarctic environments (Holzinger *et al.*, 2018). Here vegetative cells and pre-akinetes were exposed to a combination of UV-A and UV-B radiation. Surprisingly, against the authors hypothesis, vegetative cells showed a better recovery of the photosynthetic quantum yield and a higher accumulation of phenolic compounds than pre-akinetes (Holzinger *et al.*, 2018).

An important source of phenolic compounds that ward off UV could be the phenylpropanoid biosynthesis pathway, but more on this in 1.4.

Desiccation

Thinking about the transition from water to land, the first stress that comes to mind is dehydration stress, followed by desiccation stress. Low water capacities are reasonable in an environment that is more prone to dry out from time to time due to exposition to irradiance and high temperature. Therefore, desiccation tolerance is another key feature for plant terrestrialization. Here it is important to keep in mind that streptophyte algae have no mechanism for active maintenance of water homeostasis. They are poikilohydric and, as such, their water status adjusts passively to that of their environment.

When water content in the plant cells decreases, photosynthesis is directly blocked. Additionally, photosystems, especially PSII, where oxidation of water occurs, and sites for ATP generation- and carbon assimilation, are sensitive to desiccation (Allakhverdiev *et al.*, 2008). As carbon fixation processes are affected as well, stress-responsive repair mechanisms like the replacement of damaged proteins like D1 are impaired, leading to more ROS production (Müller *et al.*, 2001). Therefore, managing water content actively through stomata is a significant survival advantage of land plants. Stomata were likely present in the last common ancestor of land plants, but there were also secondary losses in some mosses and liverworts (**Publication VI**, Harris *et al.*, 2020).

Even though streptophyte algae do not possess stomata the underlying core signal cascade is still there. In embryophytes the phytohormone abscisic acid (ABA) is involved in stomata closure. This phytohormone is a major player in various phytohormone-based stress responses and has also been found in Chlorophytes and streptophyte algae (**Publication IV**; Shinozaki and Yamaguchi-Shinozaki 2003; Yoshida *et al.*, 2014). The ABA core signaling pathway involves a negative regulation cascade: ABA binds to a receptor of the PYR/PYR-1LIKE/RCAR (PYRABACTIN RESISTANCE1/PYR1-LIKE/REGULATORY COMPONENTS OF ABA RECEPTOR) family. Without the presence of ABA, these receptors inhibit the PP2C (PHOSPHATASE 2C) which in turn prevent the activity of the SnRKs (SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASEs). When ABA is bound to the proteins of the PYR/PYR-1LIKE/RCAR family, these inhibit the PP2Cs that thus no longer inhibit the SnRKs. When SnRKs get activated in the presence of ABA they can in turn activate downstream transcription factors like bZIP (basic Leucine zipper) and ion channels like SLAC (S-type anion channels), (for an overview see also **Publication IV**; Cutler *et al.*, 2010; Lind *et al.*, 2015).

Homologs for this ABA core signaling cascade are found in streptophyte algae with the PP2C/SnRK2 interaction likely emerging in Klebsormidiophyceae while the interaction with the PYL-receptor emerged at the base of Zygnematophyceae and embryophytes (Sun *et al.*, 2019; Sun *et al.*, 2020) To investigate the role of the zygnematophycean PYLs functional studies were performed on a homologous PYL in *Zygnema circumcarinatum* (ZcPYL), revealing that this PYL candidate does interact with PP2C, but this interaction is not dependent on the presence of ABA (Sun *et al.*, 2019). This however does not exclude the possibility that this cascade is involved in stress responses in Zygnematophyceae.

There are other preventive mechanisms to reduce water-loss and some streptophyte algae are shining examples. First and foremost, the class Klebsormidiophyceae stands out with a range of species being extremely tolerant to desiccation stress. The mechanisms that underpin this tolerance are diverse and on the molecular level largely obscure. Their secret can be a flexible cell wall with an alteration of the actin cytoskeleton leading to regulated cell shrinkage under desiccation stress like *in Klebsormidium crenulatum* (Holzinger *et al.*, 2011). Karsten and colleagues further showed that species from the genera *Interfilum* can protect their photosynthetic machinery under desiccation quite efficiently: in their hands, effective photosynthetic quantum yield fully recovered after rehydration (Karsten *et al.*, 2014). Similar studies were performed on the alga *Klebsormidium dissectum* which inhabits biological soil crusts in high alpine areas. Although photosynthetic performance did not recover fully after rehydration, this species was able to survive for three weeks under desiccating conditions (Karsten and Holzinger 2012).

Temperature fluctuations

In terrestrial habitats, temperature fluctuates more swiftly than in aquatic habitats. An alternation of membrane fluidity is one of the first effects fluctuations in temperature cause in a plant. Here the fluid state of the membrane influences the expression of downstream genes for cold- or heat-inducible genes, like, for example, heat-inducible genes (Horváth *et al.*, 1998). These genes in turn can facilitate the expression of heat shock proteins (HSPs), ROS scavengers that can also act as chaperon proteins to counteract protein degradation under heat conditions (e.g., Kotak *et al.*, 2007). The underlying genetic landscape of HSPs is a target for heat-stress-responsive transcription factors (HSFs) (Ohama *et al.*, 2017). The HSF – HSP network is a highly conserved molecular mechanism in prokaryotic and eukaryotic kingdoms (e.g., Richter *et al.*, 2010). In plants this process is also connected to other stresses like oxidative or high irradiance stresses, involving plastid-derived signals (for an overview see Sun and Guo, 2016). An example for plastid-derived thermotolerance could be shown by Monte and colleagues (2020).

Here, the plastid-synthesized jasmonic acid precursors OPDA and dn-OPDA (dinor-12-oxo-10,15(Z)-phytyldienoic acid) were found to also play a role in the regulation of thermotolerance genes during heat stress in land plants like *Arabidopsis* and *Marchantia* but also in *Klebsormidium nitens* (Monte *et al.*, 2020).

Besides the molecular toolkit, morphological features can also play a role in temperature tolerance. Here again the filamentous *Klebsormidium* is an exceptional example in stress tolerance in streptophyte algae. Borchhardt and Gründling-Pfaff (2020) investigated 20 *Klebsormidium* strains from Arctic and Antarctic environments and characterized those strains as psychrotolerant because of their extremophilic ability to grow in low temperatures (Borchhardt and Gründling-Pfaff, 2020). Here a defining feature is also the formation of biological soil crusts (BSCs) which protect the filamentous algae against multiple stressors (see also Karsten and Holzinger, 2012).

A few words on multifactorial stressors

In terrestrial habitats various environmental stressors can affect plant growth and survival simultaneously. The molecular ramifications of such combined stressors can have even more deleterious effects on the plant. This also holds true for temperature stress: high irradiance/ UV radiation can come in combination with heat stress, enhancing ROS production in the plastid even more. Desiccation stress can go hand in hand with freezing or with heat stress (evaporation). Additionally, heat stress combined with biotic stress can enhance the negative effects on the plant drastically, since pathogens often thrive under warmer temperatures (for an excellent overview, see also Desaint *et al.*, 2021). A fine-tuned, fast-responsive acclimation to a combined set of stressors is therefore crucial for the plant. One way to achieve this is through specialized metabolism.

1.4 Specialized plant metabolism in streptophyte algae

1.4.1 The Phenylpropanoid Pathway: The evolution of a hallmark land plant pathway for specialized metabolites

Land plant constantly adjust to fluctuating environmental conditions. One important aspect of the required responses is the production of specialized metabolites. This feature is one of the reasons why embryophytes successfully survived on land (Weng and Joseph, 2013). One of the main sources for specialized metabolism is the phenylpropanoid pathway. This pathway brings forth a diverse set of specialized metabolites which in turn can be further chemically modified (Vogt, 2010).

Phenylpropanoid-derived compounds play an important role in land-plant response to abiotic as well as biotic stressors including high irradiance, UV-radiation, fluctuations in temperature or salinity as well as pathogen attack (Dixon and Paiva, 1995; Carella *et al.*, 2019; König *et al.*, 2014; Dong and Lin, 2021). Here, phenylpropanoids, as well as metabolites from the phenylpropanoid-derived flavonoid pathway, can provide photoprotective functions while other phenylpropanoid-derived compounds like the hydrophobic polymer lignin provide structure and mechanical protection to the plant. Lignin is a hallmark land plant compound that defines vascular plant structure and growth and plays a role in a variety of stress response mechanisms like e.g., desiccation or biotic stress response (Alvarez *et al.*, 2008; Carella, *et al.*, 2019). The list of stress response mechanisms in which the phenylpropanoid pathway is directly or indirectly involved in embryophytes is long (Dixon and Paiva, 1995; Vogt, 2010).

All embryophytes make use of the pathway but the compound landscape in the embryophyte lineage in response to stressors can be highly diverse, shaped by the environment, and lineage-specific. This is also owed to the enzymatic promiscuity of the phenylpropanoid pathway, which will be further scrutinized in the discussion section of this thesis. Nevertheless, there are stress responses involving the phenylpropanoid pathway that are conserved across all embryophytes. These include the production of flavonoids when faced with UVR (Clayton *et al.*, 2018) and the activation of the pathway upon pathogen attack in the moss *Physcomitrium* (Wolf *et al.*, 2010), the liverwort *Marchantia* (Carella *et al.*, 2019,) in gymnosperms (Oliva *et al.*, 2015) and angiosperms (Dixon and Paiva 1995; Kaur *et al.*, 2010).

Interestingly, these conserved stress response strategies are likely not restricted to just land plants. Various examples in different classes of streptophyte and chlorophyte algae suggest an involvement of compounds that are conceivably phenylpropanoid-associated or -derived. Mass spectrometry-based investigations carried out during the genome project for the unicellular Zygnematophyceae *Penium margaritaceum* revealed the production of different flavonoids (Jiao *et al.*, 2020); phenolic compounds have also been detected in filamentous Zygnematophyceae, including diverse species of *Zygnema* (Holzinger *et al.*, 2018, Pichrtová *et al.*, 2013). Goiris and colleagues screened several core-chlorophytes via HPLC and could detect flavonoids like Narigenin or Kampferol (Goiris *et al.*, 2014). Lastly, a lignin-like compound was also found in three species of the Coleochaetophyceae through GC-MS as well as immunocytochemical analysis (Sørensen *et al.*, 2011).

Large enzyme-coding gene families often are responsible for land-plant characteristic biosynthetic pathways (Nelson and Werck-Reichhart, 2011; Renault *et al.*, 2017), and the underlying enzymatic setup of the phenylpropanoid pathway is no different (Vogt, 2010; Hamberger *et al.*, 2007).

Recent comparative transcriptomic and genomic analyses show that the genetic repertoire coding for the core-pathway enzymes is already present in streptophyte alga classes (Figure 5, **Publication IV**; **Publication V**; de Vries *et al.*, 2017). In the following the focus will be therefore on enzymes of the core-pathway.

The first synthesis step leading into the phenylpropanoid pathway is the conversion of Phenylalanine to cinnamate by the Phenylalanine-ammonia lyase (PAL) via non-oxidative deamination (Lois *et al.*, 1989). The enzyme is bifunctional and can also convert tyrosine (Tyrosin-ammonia lyase, PTAL) to *p*-coumaric acid and is a rare example that a single amino acid exchange can lead to a different substrate specificity (Watts *et al.*, 2006; Vogt, 2010). PALs/PTALs were long thought to have been gained via lateral gene transfer at the base of the embryophyte clade (Emiliani *et al.*, 2009). But a PAL-like enzyme could also be identified in *Klesormidium nitens* (**Publication V**; de Vries *et al.*, 2017).

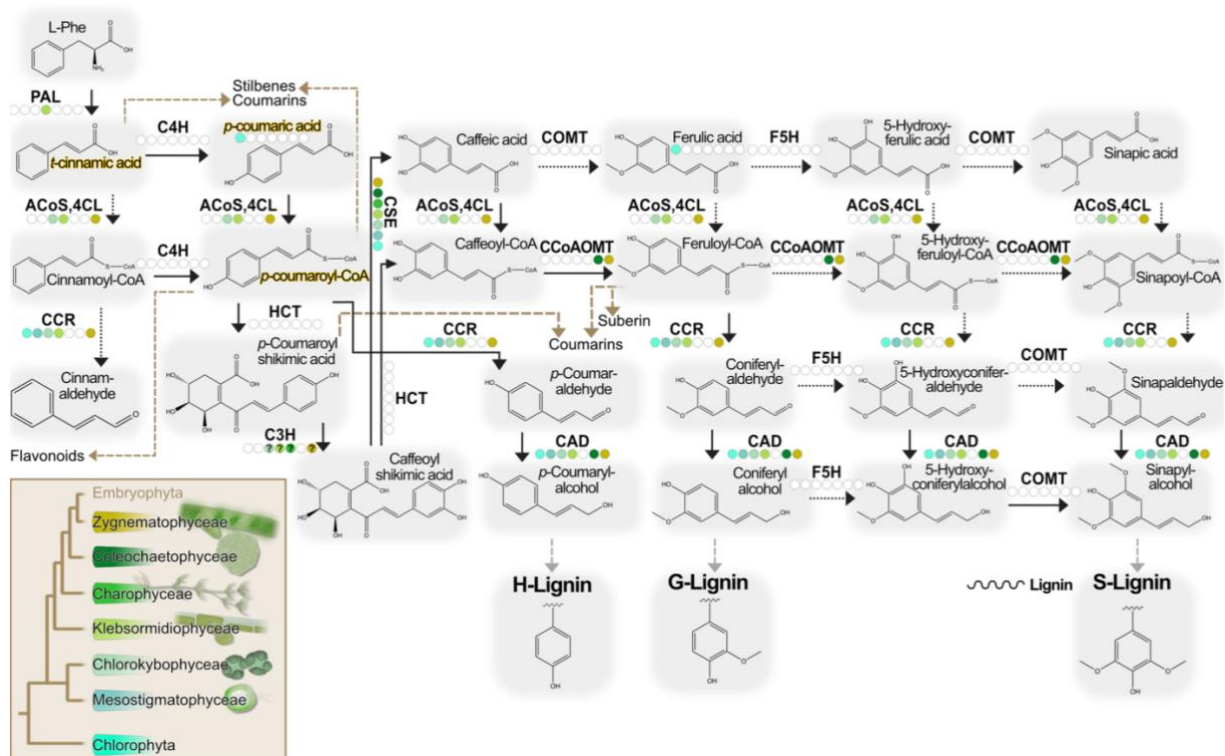


Figure 5: Schematic overview of the phenylpropanoid pathway. Structures of compounds are shown together with the corresponding enzymes. The phylogenetic evidence for these enzymes is shown by a presence/absence color code that corresponds to the lineage. A white circle means that there is no phylogenetic evidence for this enzyme in this plant lineage. Question marks indicate that the presence/absence is inconclusive (Modified from **Publication IV**).

After cinnamate is synthesized, it gets converted to *p*-coumarate by the cinnamate 4-hydroxylase (C4H) (Czichi and Kindl, 1977). For this enzyme, no clear ortholog could be found in streptophyte algae (**Publication V**; de Vries *et al.*, 2017).

It is however hypothesized that streptophyte algae and chlorophytes might share a C₄H independent route that leads to *p*-coumarate as this compound could be already detected in chlorophytes (Goiris *et al.*, 2014). The conversion of *p*-coumarate to *p*-coumeryl CoA through the AMP-forming synthase/ligase 4CL (Shockey *et al.*, 2003) is one of the key steps in the pathway. *p*-coumeryl CoA is a central compound from which diverse other pathway-derived compounds like coumarins, (iso)-flavonoids, auronones, stilbenes and lignans can emerge (Vogt 2010; Vanholme *et al.*, 2012).

1.4.2 The evolutionary conservation of lipid droplet formation

Lipid droplets (LDs) are cytosolic compartments that can form under various stress conditions and are present throughout the eukaryotic kingdom. Yet their evolution throughout the green lineage and their role during changing environmental conditions, especially regarding terrestrialization events still needs further investigation. Cytosolic lipid droplets form at the endoplasmic reticulum, consist mainly of neutral lipids like triacylglycerols (TAGs) and sterol esters (SE) and are surrounded by a phospholipid monolayer (Pyc *et al.*, 2017b). LD formation has been extensively studied in seeds, where they act as storage reservoirs (Gasulla *et al.*, 2013). The flowering plant *Arabidopsis thaliana* has been used as a model to study LD formation and their differential accumulation. Here, LD abundance was found to increase in connection to abiotic stressors like heat -, cold- and drought stress (Higashi *et al.*, 2015; Gasulla *et al.*, 2013).

Especially drought stress and nitrogen-deficient environments are major triggers for LD formation not just in land plants (e.g., Coulon *et al.*, 2020) but in algae as well (Breuer *et al.*, 2012; Jia *et al.*, 2015). In the last decade, there has been considerable interest in lipid-producing chlorophytes, especially owing to promising biotechnological applications. In this context, many analyses involve nitrogen starvation to enhance TAG synthesis and use it for biofuel production (Yang *et al.*, 2013; Valledor *et al.*, 2014). In the model chlorophyte alga *Chlamydomonas*, LD formation was even investigated in iron-deficient environments (Devadasu and Subramanyam, 2021). There it was shown that nitrogen starvation and iron deficiency both trigger a remodeling of the chloroplast membrane lipids to form TAG (Devadasu and Subramanyam, 2021; Yang and Benning, 2018). Yet, chlorophyte algae lack many of the genes that code for the protein chassis that makes up the LDs of land plants. This begs the question of when the embryophytic chassis for LD formation emerged. Here, streptophyte algae will harbor important information.

Even though there are fewer reports on this molecular response in streptophyte algae compared to the biotechnological promising chlorophytes, it has been shown that the genetic repertoire for some hallmark Lipid droplet related proteins is already present in different classes of streptophyte algae (de Vries and Ischebeck, 2020). Among those are gene homologs coding for

proteins like LDAP (LD ASSOCIATED PROTEIN), CAS (CYCLOARTENOL SYNTHASE), OLE 7 (OLEOSIN) or HSD1 (STEROLEOSIN). These proteins are involved in lipid droplet formation in *Arabidopsis* (Pyc *et al.*, 2017a, see also Ischebeck *et al.*, 2020). The studies discussed in this thesis provide further insights into the underlying lipid-droplet protein-coding genetic landscape in Zygnematophyceae (**Publication II, Publication III, Preprint VII**).

Morphologically lipid droplet like structures have often been observed in different classes of streptophyte algae, often dubbed as “storage granules” or “lipid bodies” (e.g., Holzinger *et al.*, 2011; Holzinger *et al.*, 2018). But what is often lacking is the confirmation that the morphological structures observed are indeed land plant-like lipid droplets, characterized by their associated proteome (Lipid-droplet associated proteins) (Brocard *et al.*, 2017), as well as their neutral lipid content. Summing up, lipid droplet formation is a promising feature of conserved stress response which opens up intriguing questions about a potential role during plant terrestrialization and deserves further investigation in streptophyte algae.

2 Aim of this Thesis

As much as plant terrestrialization is one of the most important events in the evolution of the green lineage, there are still many mysteries surrounding this event. The overall looming question that is frequently asked in that context is: why was plant terrestrialization a singular event? One attempt to answer this question is to look at the environmental challenges that the first land plants likely faced and how they overcame different stress conditions. This boils down to a new question: what molecular traits aided the first land plants in their conquest of land? To understand this, we need to compare the molecular chassis for stress responses of land plants and their algal relatives (Figure 6). In this thesis this approach is pursued through the investigation of streptophyte algal stress response. The goal is to unravel parts of the evolutionary conserved, fine-tuned stress responsive toolkit that aided the earliest land plants in the adaptation to their new environment — a core toolkit that is shared by streptophyte algae and land plants.

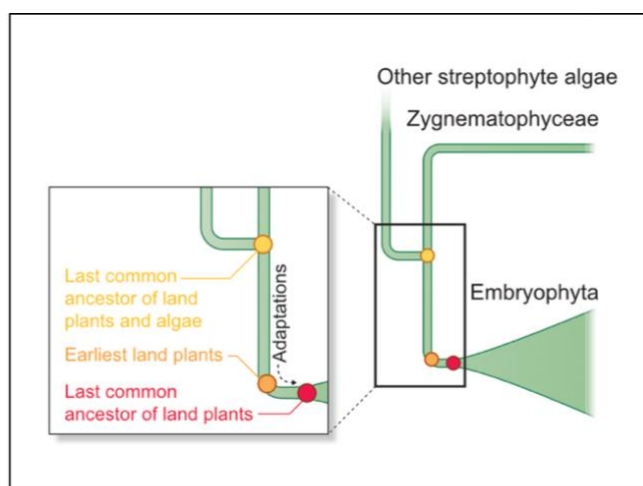


Figure 6: Reconstructing the biology of the earliest land plant. A comparison of traits of the last common ancestor of land plants and algae (yellow dot) and the last common ancestor of land plants (red dot) might shed light on the traits of the earliest land plant. Adaptions that were gained on land must be subtracted. (Modified from **Publication I**).

This thesis is based on two approaches (Figure 7). The first approach, covered in **Chapter I (i)** is to look at abiotic stress response in streptophyte algae on a broader scale. Compared to information on land plant stress response, there is still a lot unknown about the molecular stress response kit in streptophyte algae. Providing a broader overview of which stress response mechanisms are present is therefore more reasonable. Here a treatment with abiotic stressors was chosen which were applied to different species of the Zygnematophyceae (*Mougeotia sp.*, *Mesotaenium endlicherianum*). These algae were then analyzed based on morphology and photophysiology.

The overall focus here however was on the landscape of their transcriptomic response. This was done using untargeted analyses that aim to find shared core-stress mechanisms through comparative transcriptomics.

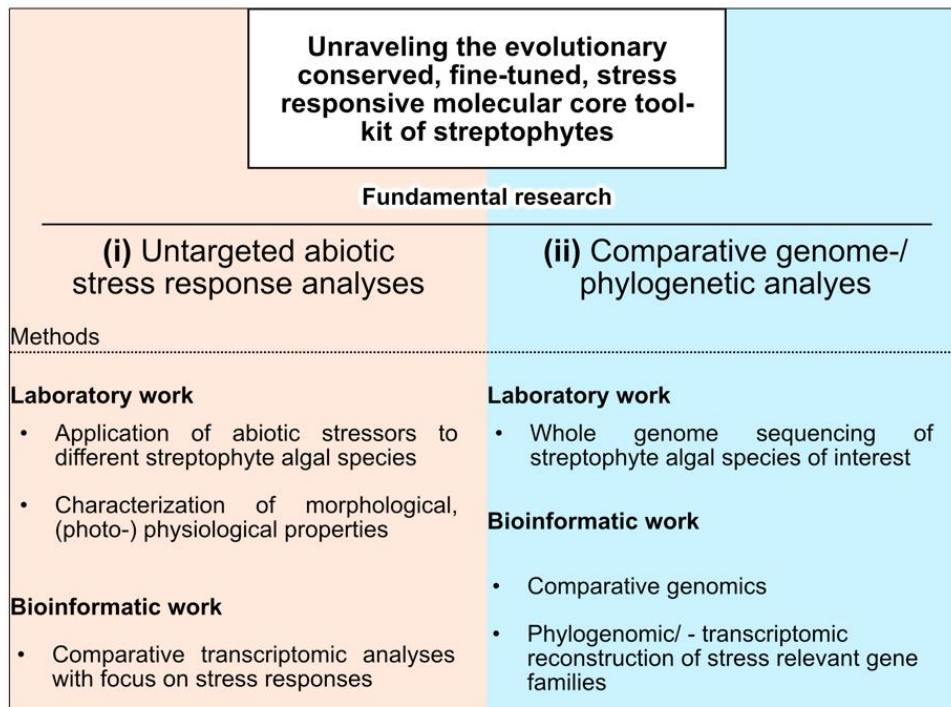


Figure 7: Two different approaches to investigate the hypothesized core-stress responsive toolkit streptophyte algae and land plants share. The first approach on the left **(i)** comprises untargeted abiotic stress response analyses. The second approach on the right **(ii)** describes comparative genome-/phylogenetic analyses. Both approaches are top-down approaches. To be able to infer a shared molecular toolkit it is for both approaches crucial to always compare with already known land plant data. Note that both approaches fall under the category of fundamental research and therefore it is the aim to advance knowledge on this still narrowly researched topic of streptophyte algae stress response.

The second approach, covered in **Chapter II (ii)**, focuses on comparative and evolutionary genomics. In recent years the unraveling of streptophyte alga genomes steadily advanced (see **Table 1**). Providing *de-novo* assembled genomes can help greatly with in-depth transcriptomic stress response analyses on promising streptophyte alga species. Furthermore, a phylogenetic approach can shed light on whole gene-families that are well studied in land plants and are involved in a variety of stress response mechanisms inside the plant cell but are barely known in streptophyte algae. This approach can help to reconstruct when certain traits emerged during streptophyte evolution and if those molecular traits possibly aided in terrestrialization. Both approaches can eventually pave the way for in-depth analyses of specific stress response mechanisms in streptophyte algae.

3 Chapter I: A broad view of abiotic stress response in streptophyte algae

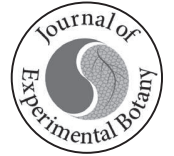
3.1 **Publication I:** Evo-physio: on stress responses and the earliest land plants

This review paper was published online in the Journal “Journal of Experimental Botany” in January 2020. The full article can also be found online:

<https://doi.org/10.1093/jxb/eraa007>

Contribution of Janine Fürst-Jansen, first author

J. M. R. Fürst-Jansen planned and wrote the chapter “is the physiology of streptophytes unique?” and critically read and revised the manuscript.



REVIEW PAPER

Evo-physio: on stress responses and the earliest land plants

Janine M.R. Fürst-Jansen¹, Sophie de Vries², and Jan de Vries^{1,3,*},

¹ University of Göttingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, D-37077 Göttingen, Germany

² Population Genetics, Heinrich-Heine University Düsseldorf, Universitätsstr. 1, D-40225 Düsseldorf, Germany

³ University of Göttingen, Göttingen Center for Molecular Biosciences (GZMB), D-37077 Göttingen, Germany

* Correspondence: devries.jan@uni-goettingen.de

Received 11 October 2019; Editorial decision 6 January 2020; Accepted 7 January 2020

Editor: Henrik Buschmann, University of Osnabrück, Germany

Abstract

Embryophytes (land plants) can be found in almost any habitat on the Earth's surface. All of this ecologically diverse embryophytic flora arose from algae through a singular evolutionary event. Traits that were, by their nature, indispensable for the singular conquest of land by plants were those that are key for overcoming terrestrial stressors. Not surprisingly, the biology of land plant cells is shaped by a core signaling network that connects environmental cues, such as stressors, to the appropriate responses—which, thus, modulate growth and physiology. When did this network emerge? Was it already present when plant terrestrialization was in its infancy? A comparative approach between land plants and their algal relatives, the streptophyte algae, allows us to tackle such questions and resolve parts of the biology of the earliest land plants. Exploring the biology of the earliest land plants might shed light on exactly how they overcame the challenges of terrestrialization. Here, we outline the approaches and rationale underlying comparative analyses towards inferring the genetic toolkit for the stress response that aided the earliest land plants in their conquest of land.

Keywords: Charophytes, earliest land plants, exaptations, plant evolution, plant terrestrialization, streptophyte algae, stress physiology, terrestrial algae.

Introduction

Green evolution: from the origin of photosynthetic eukaryotes to the earliest land plants

Photosynthetic eukaryotes probably first emerged >1.5 billion years ago (Butterfield, 2000; Eme *et al.*, 2014; Bengtson *et al.*, 2017). Underlying the origin of photosynthetic eukaryotes was the endosymbiotic uptake of a free-living cyanobacterium by a heterotrophic protist—an event that gave rise to the Archaeplastida (reviewed by Archibald, 2015; Martin *et al.*, 2015; de Vries and Gould, 2018). There are three types of

Archaeplastida: the red algae (rhodophytes), the glaucophytes, and the green lineage (Keeling, 2013; Archibald, 2015; Jackson *et al.*, 2015). The green organisms make up the Chloroplastida (Fig. 1)—a name that should be given preference over the previous label for that clade, Viridiplantae (Adl *et al.*, 2005, 2019). Within the Chloroplastida, we find both green algae and the land plants (reviewed by Leliaert *et al.*, 2012; de Vries *et al.*, 2016).

The green lineage separated roughly a billion years ago into the chlorophytes and the streptophytes (Zimmer *et al.*, 2007; Parfrey *et al.*, 2011; Morris *et al.*, 2018). While the chlorophytes are generally perceived as the clade comprising famous

green algae (such as *Volvox*, *Ulva*, and *Chlamydomonas*), the streptophytes are best known as the clade containing the land plants. However, there is more to the lineage of streptophytes. In the phylogeny of streptophytes sits—next to the land plants—the paraphylum of streptophyte algae (Fig. 1). It is this grade that one must turn to in order to understand the origin of land plants.

All land plants evolved from a single streptophyte algal progenitor (reviewed in de Vries and Archibald, 2018). The streptophyte algae are a group of mainly freshwater and terrestrial algae—with a few representatives living in brackish environments (Lewis and McCourt, 2004; Becker and Marin, 2009; Fig. 1). That the streptophyte algal ancestors of land plants lived in freshwater—as opposed to marine—environments is considered a major factor leading to terrestrialization: starting from a freshwater environment such as a pond, there was smooth

passage along the hydrological gradient towards land (see discussions in Becker and Marin, 2009; Delwiche and Cooper, 2015; de Vries and Archibald, 2018). It is this stepwise conquest of land along the hydrological gradient where the earliest land plants—or the first common ancestors of land plants—are evolutionarily and ecologically situated (Figs 1, 2). This ecological setting, however, does not constitute a reason for the singular global conquest of land by streptophytes. Indeed, photosynthetic eukaryotes might have had a freshwater origin (Delwiche and Cooper, 2015; de Vries and Archibald, 2017; Lewis, 2017; Ponce-Toledo *et al.*, 2017; Sánchez-Baracaldo *et al.*, 2017). Streptophytes are also not the only photosynthetic eukaryotes that dwell in freshwater environments and on land. Various other algae, including chlorophytes (for an overview, see Holzinger and Karsten, 2013), diatoms (for an overview, see Souffreau *et al.*, 2013), red algae (e.g. *Porphyridium*, see John,

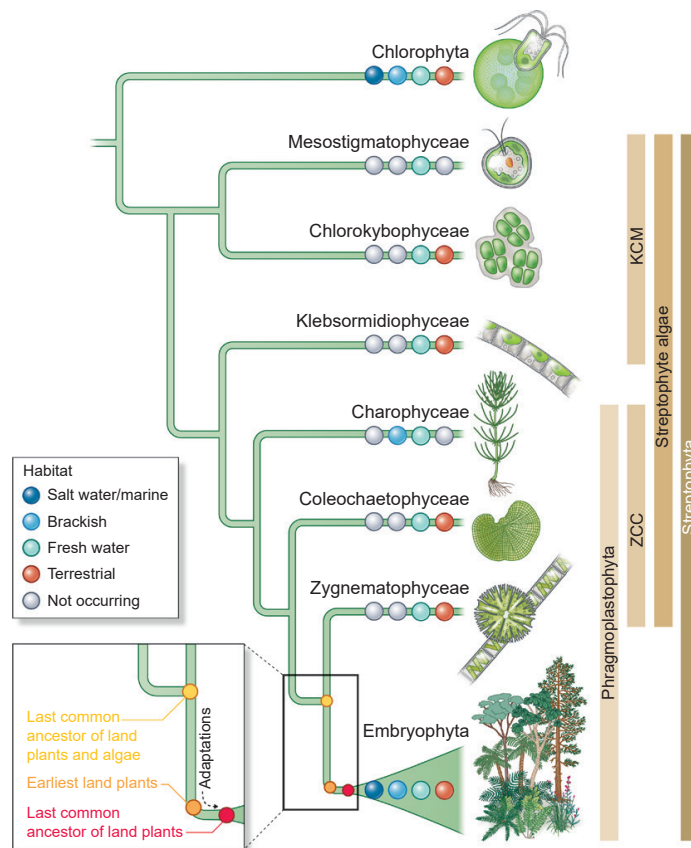


Fig. 1. Terrestrial organisms are found across the green lineage. A cladogram shows the deep split of the green lineage into the clades Chlorophyta and Streptophyta. The Streptophyta are composed of the paraphylum streptophyte algae and the monophyletic Embryophyta (land plants). Streptophyte algae can be broken up into the paraphyla KCM (for Klebsormidiophyceae, Chlorokybophyceae, and Mesostigmatophyceae) and ZCC (for Zygnematophyceae, Coleochaetophyceae, and Charophyceae; de Vries *et al.*, 2016). ZCC streptophyte algae and land plants form the monophyletic clade Phragmoplastophyta. Taken in their entirety, Chlorophyta occur in habitats ranging from marine saltwater, to freshwater, to terrestrial (row of dots). Streptophyte algae mainly occur in freshwater and terrestrial environments; some Charophyceae live in a brackish habitat. While the Embryophyta are mainly terrestrial, some have secondarily moved back to a freshwater habitat; some have even conquered a new habitat: saltwater (e.g. sea grasses). Inset: the Zygnematophyceae are the closest algal relatives of land plants and they hence share with the clade of Embryophyta the last common ancestor of land plants and algae (yellow dot); along the trajectory from that last common ancestor of land plants and algae (yellow dot) to the last common ancestor of land plants (red dot) are the earliest land plants to be found (orange dot). Inferring the biology of the earliest land plants requires a subtraction of the traits ("adaptations") that were gained on land, that is en route to the last common ancestor of land plants (from the orange to the red dot; see also Fig. 2).

1942), and many more, are terrestrial, too (Hoffmann, 1989). Hence, the question is not only what allowed for the origin of land plants but also what allowed for their unique success—a success that resulted in the global conquest of land. In this review, we will explore the complexities underlying these questions and make a case for dissecting one—which is by far not the only—key aspect of the biology of the earliest land plants: adequately responding to terrestrial stressors.

Synapomorphies and the global success of land plants

Embryophytes (land plants) are defined by a series of traits. For example, land plants undergo a cycle where they alternate between a diploid sporophyte and a haploid gametophyte. An alternation of generations that involve two multicellular generations of different ploidy is one of the features of land plants that sets them apart from their algal ancestors (for more on

this topic, see Bowman *et al.*, 2016; Horst *et al.*, 2016; Rensing, 2018). Among such embryophytic traits, we also might find those features that allowed for the success of the monophylum of land plants.

As the name implies, having embryos and embryogenesis is a signature feature of embryophytes. Broadly speaking, the embryo is a parentally supported complex structure with different tissue types. The exact organization and structure of embryos varies across the diversity of land plants concomitant with the dominance of sporophyte and gametophyte (for an overview of the underlying process, see Rensing, 2016). Support both through the parent organism and the structural framework that the embryo defines is thought to have been adaptive for living on land: the pre-defined structure of the embryo leads to an upright body plan with specialized tissues, both of which can foster nutrient uptake and nuanced responses to environmental cues via specialized cells (Rensing, 2016). When pondering the evolution of plant embryos, the seed of spermatophytes comes

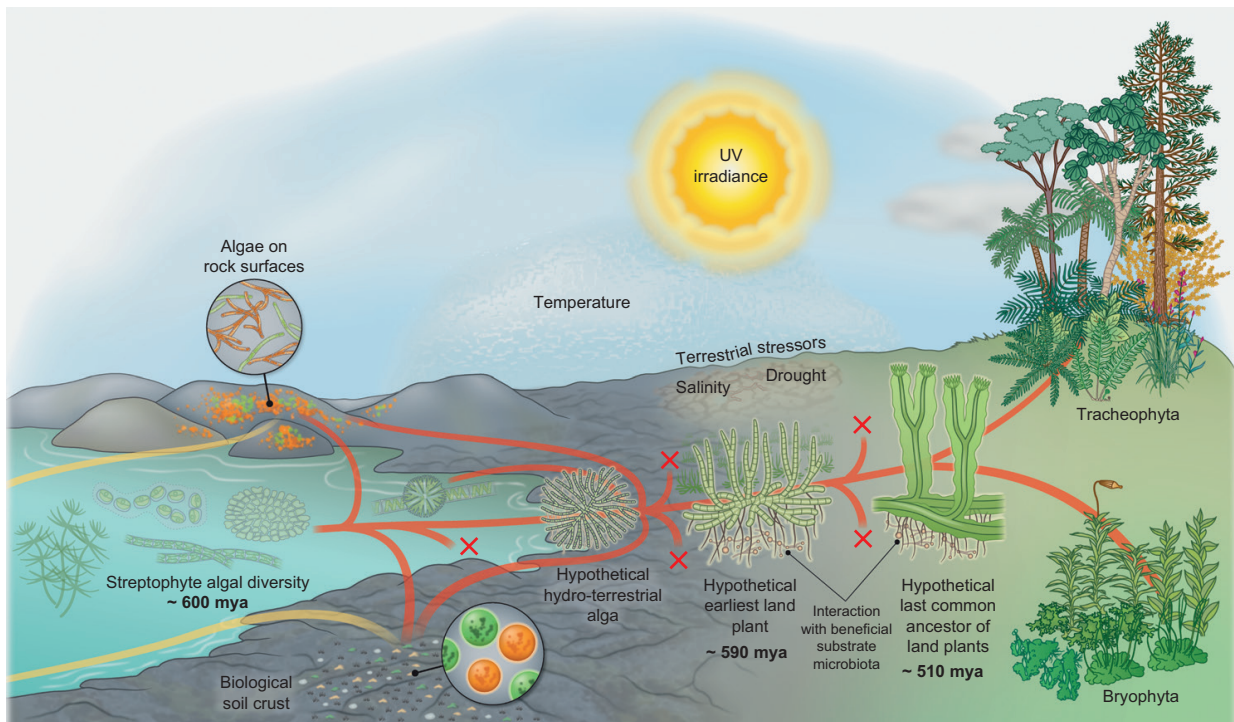


Fig. 2. The earliest land plants: an evolutionary scenario for the conquest of land by streptophytes. Streptophyte algae are the only photosynthetic eukaryotes from which the macroscopic land flora evolved (red lines). That said, throughout the course of evolution, algae from various other lineages have colonized land (yellow lines)—but also streptophyte algae have continuously and independently made the wet to dry transition (convergence of red and yellow). Throughout history, numerous lineages have become extinct ('X' labels). Terrestrial algae of various taxonomic affiliations dwell on rock surfaces and form biological soil crusts. From the diversity of the paraphyletic streptophyte algae, however, did an organism whose descendants eventually conquered land on a global scale emerge: a likely branched filamentous—or even parenchymatous—organism that formed rhizoidal structures and experienced desiccation from time to time. From this 'hypothetical hydro-terrestrial alga', the lineages of Zygnematophyceae and embryophytes (land plants) arose. In its infancy, the trajectory leading to the embryophytes was represented by the—now extinct (see also Delaux *et al.*, 2019)—earliest land plants. The earliest land plants probably interacted with beneficial substrate microbiota that aided them in obtaining nutrients from their substrate. Furthermore, the earliest land plants had to successfully overcome a barrage of terrestrial stressors (including UV and photosynthetically active irradiance, drought, drastic temperature shifts, etc.). They succeeded because they had the right set of traits—a mix of adaptations that were selected for in their hydro-terrestrial algal ancestors, exaptations, and the potential for co-option of a fortuitous set of genes and pathways. During the course of evolution, some members of the populations of the earliest land plants gained traits that are adaptive in terrestrial environments (such as some form of water conductance, stomata-like structures, embryos, etc.); eventually, the 'hypothetical last common ancestor of land plants' emerged. From this ancestor, the extant bryophytes and tracheophytes evolved. While the exact trait repertoire of the hypothetical last common ancestor of land plants is uncertain, it will certainly have entailed properties of vascular and non-vascular plants. What is also certain is that the last common ancestor of land plants had traits of algal ancestry. (All dating is roughly based on Morris *et al.* [2018])

to mind. The seed is a structure that is highly resistant to terrestrial stressors (principally desiccation—see stimulating discussions in [Oliver et al., 2000](#)). Seeds obtain their stress resistance through molecular mechanisms such as the accumulation of LATE EMBRYOGENESIS ABUNDANT (LEA; e.g. [Dure et al., 1981](#); [Xu et al., 1996](#)) proteins and seed dormancy regulation via abscisic acid (ABA; reviewed in [Holdsworth et al., 2008](#)); more on this below. While the seed clearly is a derived structure of spermatophytes, the molecular framework that underpins their stress resilience is probably conserved across the breadth of land plant diversity ([Cuming et al., 2007](#); [Eklund et al., 2018](#)). This framework was hence likely to have been present in the last common ancestor of land plants—and potentially even before.

Land plants have evolved a number of complex structures that are adaptive in a terrestrial habitat ([Harrison, 2017](#)). Among these structures are stomata. There is much debate about the exact trajectory of stomata evolution; for an overview, see [Chater et al. \(2017\)](#). Yet, it is considered probable that some sort of stomata were a feature of the last common ancestor of land plants (e.g. [Duckett and Pressel, 2018](#)). The ancestral function of stomata is, however, ambiguous and much debated (see, for example, [Duckett and Pressel, 2018](#); [Pressel et al., 2018](#)). A similar case applies to water-conducting tissues. Land plants need to allocate water from their substrate, which is facilitated by rooting structures that range from rhizoids in non-vascular plants to the ‘true roots’ of vascular plants (reviewed by [Jones and Dolan, 2012](#); [Kenrick and Strullu-Derrien, 2014](#); [Hetherington and Dolan, 2017](#)). Root-mediated water conduction from the substrate through the entire plant is a textbook process of vascular plants that is clearly adaptive in aeroterrestrial environments. Some mosses and liverworts have water-conducting tissues such as hydroids—yet these probably evolved multiple times independently ([Ligrone et al., 2000](#)) and most do not have water-conducting tissues. Surprisingly, [Xu et al. \(2014\)](#) highlighted that the same group of transcription factors (TFs), the NACs, that regulate xylem differentiation in the vascular plant *Arabidopsis thaliana* also regulate hydroid differentiation in the moss *Physcomitrella patens*. This has potential implications for vascular or non-vascular water-conducting cells in the last common ancestor of land plants (see [Fig. 2](#)). The findings of [Xu et al. \(2014\)](#) thus underscore the genetic capacity for the earliest land plants to have gained a complex system for water conduction.

This list could be continued but there is a stumbling block to most of the above-named traits: they define land plants as we know them today. These traits were likely to have been present in the last common ancestor of all land plants. Yet, it is difficult to put the gain of these traits into the right order that might enable us to reconstruct a scenario for the origin of land plants. Here the closest streptophyte algal relatives of land plants can help. Among streptophyte algae, we find (i) traits that were once classified as land plant specific and (ii) genes that are required for realizing such traits—even if they are not fully realized in the algae or used in an entirely different manner as in land plants.

Inferring trait evolution towards understanding the singularity of plant terrestrialization

The last common ancestor of land plants was an embryophyte. As elaborated in the previous paragraph, this common ancestor must have had an array of the synapomorphic traits that define embryophytes. These traits give a post-hoc perspective on the singularity of the origin of the embryophytic clade: The last common ancestor was probably already established on land. However, when did the decisive traits evolve if we consider the earliest land plants and, thus, the organisms that conquered land (see [Fig. 2](#))? To pinpoint those features that might have allowed for the conquest of land, we have to look at what happened before the last common ancestor of land plants lived—we have to resolve features of the biology of the earliest land plants ([Figs 1, 2](#)). To do so, we need to add an informative set of streptophyte algae to the picture.

In the past few years, garnering phylogenetic (e.g. [Wickett et al., 2014](#); [Puttick et al., 2018](#); [Leebens-Mack et al., 2019](#)) and sequence data (e.g. [Hori et al., 2014](#); [Delaux et al., 2015](#); [Ju et al., 2015](#); [de Vries et al., 2018a](#); [Nishiyama et al., 2018](#)) for streptophytes has gained momentum, resulting in a changing picture of early land plant evolution. In light of these new data, the notion that the embryophyte lineage might be split into two monophyletic groups ([Puttick et al., 2018](#)) raises the question of what the properties of the common ancestor of embryophytes might be; its repertoire of traits could have entailed features of both bryophytes (mosses, liverworts, and hornworts) and tracheophytes (vascular plants). Here, a streptophyte algal perspective will help, too. Disentangling the transition from the earliest common ancestor to the last common ancestor of land plants (see also [Fig. 2](#)) will not only illuminate the properties present and relevant during the earliest steps of plants on land, but also those at the base of the land plants.

In the previous section, we have listed a number of adaptations of embryophytes to living on land. If we projected their origin onto the trajectory from streptophyte algal ancestor to extant embryophytes, we would find that the streptophyte algal progenitor probably possessed a few of these key traits—at least in a rudimentary fashion. Palpably, these include rhizoids and multicellular growth.

Rhizoids or similar structures can be found in all phragmoplastophytic streptophyte algae. The most obvious cases of these are the multicellular rhizoids of Charophyceae, whose statoliths are even involved in the modulation of gravitropism (e.g. [Leitz et al., 1995](#)). The Coleochaetophyceae do not form rhizoids, but have special hairs that rest in a sheath, whose possible homology with rhizoids was discussed by [Graham et al. \(2012\)](#). Among the Zygnematophyceae, which are phylogenetically most closely related to land plants, rhizoid formation for providing anchorage to a substrate has been meticulously described in *Spirogyra* ([Yoshida et al., 2003](#); [Ikegaya et al., 2008](#); [Yoshida and Shimmen, 2009](#)). Rhizoid formation is, hence, a probable feature of the earliest land plants ([Fig. 2](#)).

Multicellularity in phragmoplastophytic streptophyte algae can be supported by a simplex meristem/apical cell structure.

Charophyceae have an apical cell at the tip of the shoot-like structure that confers its erect growth (Pickett-Heaps, 1967). In *Coleochaete*, varying growth morphologies occur that range from branched filaments to discoidal parenchymatous growth (for an overview, see Delwiche *et al.*, 2002); morphogenesis of the latter is underpinned by meristems (Dupuy *et al.*, 2010). During the course of evolution, the zygnematophyceae body plan probably experienced some reduction—as, for example, seen in the desmids that reverted to unicellularity. However, as already indicated previously, some filamentous Zygnematophyceae have a few morphological surprises up their sleeves, including the formation of rhizoidal holdfasts (e.g. Ikegaya *et al.*, 2008) and branching (e.g. Stancheva *et al.*, 2014). Altogether, this suggests that the earliest land plants probably had a body plan that entailed at least branching filaments—if not parenchymatous growth—likely to resemble to some degree the body plans found among the diversity of the genus *Coleochaete* (Fig. 2; Delwiche and Cooper, 2015). A multicellular body plan facilitates the differentiation of cells that can specialize in the responses to environmental stressors, for example the uppermost layer of a hypothetical body plan being particularly rich in compounds that act as sunscreens. Thus, multicellularity might have provided selective advantages for the earliest land plants when facing stress on *terra firma*.

Genetic potential and the evolution of decisive traits

The embryophytic transcriptionally active protein (TAP) repertoire has often been proposed to explain the complex developmental phenotypes and high plasticity in environmental responses of land plants (Lang *et al.*, 2010; Mähönen *et al.*, 2014; Harrison, 2017; Scheres and van der Putten, 2017; Wilhelmsson *et al.*, 2017). Recent studies (Catarino *et al.*, 2016; Wilhelmsson *et al.*, 2017), however, showed that >80% of the TAP family repertoire of land plants is present in streptophyte algal genomes and transcriptomes. It may therefore be assumed that an interconnected network of TAPs and their downstream targets already fine-tuned the biology of the earliest land plants. Importantly, the TAP repertoire does more than the actualization of specific traits: TAPs create a fertile ground for genetic—and hence evolutionary—innovation through a few minor modifications such as spatial and temporal modification of the expression of a whole cascade. Thus, small changes in TF number, expression, or binding capacity can have profound effects on biological phenotypes. For example, the LOTUS JAPONICUS ROOTHAIRLESS LIKE (LRL) family in *A. thaliana* is separated into two antagonistically acting classes that have strong tissue-specific expression (Breuninger *et al.*, 2016). The antagonistic network in *Arabidopsis* gave rise to a control mechanism of a variety of morphological developments, while in *Marchantia polymorpha* one single copy gene regulates the development of rhizoids and thalli (Breuninger *et al.*, 2016). A directly stress-related example for the land plant-like TAP repertoire of streptophyte algae might be the GRAS TFs. GRAS TFs regulate both development and stress response of land plants (Hirsch and Oldroyd, 2009; B. Zhang

et al., 2018). In the recently reported genome sequence of the two Zygnematophyceae *Spirogloea muscicola* and *Mesotaenium endlicherianum* (Cheng *et al.*, 2019), the authors found that already the ancestor of Zygnematophyceae and land plants shared an expanded repertoire of genes coding for GRAS TFs. Some of these GRAS TFs are thus strong candidates for regulating conserved stress responses. Altogether, TAPs offer a plausible explanation for the rapid diversification in form and function (including stress-related) that we see across the diversity of land plants; the potential for this TAP-based radiation of form and function was already present in the algal relatives of land plants.

A tangible example of the fertile ground that the genetic material of streptophyte algae offers is the genetic toolkit to establish an interaction with symbiotic fungi (Delaux *et al.*, 2015). Most (>70%) of the extant diversity of embryophytes engage in symbiosis with arbuscular mycorrhiza that aid in obtaining nutrients from the substrate (Delaux *et al.*, 2014; Field and Pressel, 2018). Land plants use a core signaling toolkit to establish such symbioses (Parniske, 2008). Symbioses with beneficial substrate microbiota are thought to have aided the earliest land plants in gaining a foothold in the terrestrial habitat (Delaux *et al.*, 2013; Field *et al.*, 2015), and orthologous genes that are involved in upstream signaling processes required for the establishment of arbuscular mycorrhizal symbiosis in land plants are present in streptophyte algae (Delaux *et al.*, 2015). Indeed, zygnematophyceae orthologs of calcium- and calmodulin-dependent protein kinase (CCaMK; a key component of the symbiosis toolkit) rescued *Medicago* mutants defective in this gene and hence mycorrhization (Delaux *et al.*, 2015). Yet, the genetic framework for the downstream signaling has only emerged after diversification of the respective gene families in the ancestor of land plants (Delaux *et al.*, 2015). Thus, building on a conserved chain of upstream signaling, elaborate developmental processes could emerge.

Symbionts are recognized by plants through receptor-like kinases (RLKs), a major family among which are the LysMs (reviewed by Oldroyd, 2013). Pivotal to the establishment of symbioses is the recognition of lipochitooligosaccharides secreted by arbuscular mycorrhizae via specific LysMs (Oldroyd, 2013; Sun *et al.*, 2015). However, LysMs and other RLKs do not only play a role in formation of arbuscular mycorrhizae but are general players during symbiotic and pathogenic interactions. Fabacean LysM receptors are involved in the recognition of lipochitooligosaccharides of rhizobia to establish colonization and nodulation of their hosts' roots (Limpens *et al.*, 2003; Madsen *et al.*, 2003; Radutoiu *et al.*, 2003; Arrighi *et al.*, 2006; Smit *et al.*, 2007). In addition to their role in symbiotic interactions, different RLKs, including LysMs, are also the gatekeepers of pathogen responses in angiosperms, recognizing conserved molecular patterns of microbes, such as flagellin or chitin (Gómez-Gómez and Boller, 2000; Kaku *et al.*, 2006; Miya *et al.*, 2007). While we have little insight into RLK function in pathogen recognition outside of angiosperms, important insights have come from the moss *Physcomitrella patens*. *Physcomitrella patens* is able to sense and respond to chitin and encodes at least one functional chitin sensing CERK1 homolog (Bressendorff *et al.*, 2016). CERK1 is thus one of the few LysM-RLKs that was probably involved in plant immune

signaling (upon being challenged by fungi) in the last common ancestor of all land plants.

The genome of *Chara braunii* provided some insight into the evolutionary history of LysM receptors (Nishiyama *et al.*, 2018). It appears that the common ancestor of *C. braunii* and land plants may have had a single LysM member and that each lineage has undergone their specific family expansions. Without functional studies, it is thus equally plausible that LysMs of streptophyte algae function in response to pathogens or symbionts—or both; hence, the same ambiguity currently applies to inferences of LysM function in the earliest land plants. What we know, however, is that streptophyte algae have the genetic toolkit to sense microbial associates, mutualistic or pathogenic microorganisms. Streptophyte algae associate with an entire microbiome of fungi and bacteria (Knack *et al.*, 2015) and we can predict, based on the presence of LysMs, that they are able to recognize and respond to these microbes in some manner. A similar prediction may be made for the earliest land plants.

It was argued that fungi evolved the ability to degrade plant material before land plants came to be (Berbee *et al.*, 2017). This suggests, as one would expect, that not all of the earliest land plants' microbiome was friendly; they might have been assaulted by foes that their ancestor already faced in freshwater habitats. Another type of pathogen receptor is shared by streptophyte algae and land plants: the intracellular resistance (*R*) genes that encode proteins with nucleotide-binding domain and leucine-rich repeat (NBS-LRR) domains (Gao *et al.*, 2018; Han, 2019). Yet, again, streptophyte algae seem to have evolved different domain associations from land plants and only a few land plant-like proteins of the toll interleukin 1 receptor (TIR)-NBS-LRR class are present in some streptophyte algae (Gao *et al.*, 2018; Han, 2019).

Only through detailed studies of the interactions between streptophyte algae and their microbiome (see Delaux *et al.*, 2015; Knack *et al.*, 2015; Gao *et al.*, 2018; reviewed in de Vries *et al.*, 2018b; Han, 2019) will we be able to infer the plant-microbe interaction toolkit of the earliest land plants. Yet as to the exact components that the earliest land plants used, the power of inference based on extant systems might be limited: receptor families have undergone many lineage-specific expansions and reductions, and are co-opted for their specific environment, which—most importantly—is co-evolving with their hosts. In contrast, the evolution of the responses to abiotic stressors may be reconstructed more easily, as the abiotic environment—albeit dynamically changing—is not shaped by an evolutionary arms race with other organisms. This touches upon a series of traits that must have been present in the earliest land plants and enabled them to overcome abiotic terrestrial stressors.

Streptophytic stress signaling: the evolution of an essential prerequisite for the conquest of land

Dwelling on *terra firma* comes with various challenges. Foremost among these are various abiotic stressors, including drought and desiccation, UV irradiation, and rapid changes

in temperature—but also changes in substrate quality such as pH, salinity, and nutrient variation. Land plants have evolved an elaborate stress response framework. This framework includes the perception of stressors, signal transduction involving ubiquitous molecules such as reactive oxygen species (ROS) and specific phytohormones such as ABA, and finally the appropriate adjustment of the physiology of the plant cell—all of which further hinges on the duration of a stress (Kranmer *et al.*, 2010). The connection between stress input and adjustment is mediated by the plant perceptron. The plant perceptron is a layered network of input signals converging in signal transduction pathways that target regulators of plant growth and physiology (Scheres and van der Putten, 2017). This response system can be considered as an additional defining trait of embryophytes. It is conceivable that the roots of this trait run deeper—already the earliest land plants had to successfully overcome the challenges that the terrestrial stressors posed in order to first colonize land and then radiate on it.

Streptophyte algae are now known to have genes for stress response that were previously thought to be characteristic for land plants. While the functions of many of these genes have not yet been tested, their mere presence warrants attention. These genes represent the ancestral gene pool from which the embryophytic genes with functions in stress response have evolved. When exactly the embryophytic function (if there is a unique function among all embryophytes) evolved can only be identified by a combination of comparative functional and bioinformatic approaches across and outside the monophyletic land plants. Independent of the function they hold nowadays in streptophyte algae, it is these genes that are prime candidates for being part of the 'terrestrialization toolkit'. The idea is that this toolkit entailed adaptive and exaptive genes that provided a selective advantage during plant terrestrialization (see also de Vries and Archibald, 2018). Standing out among these candidates is the notion that streptophyte algae have the genetic capacity to utilize phytohormone-based signaling pathways (Delaux *et al.*, 2012; Hori *et al.*, 2014; Ju *et al.*, 2015; Van de Poel *et al.*, 2016; Ohtaka *et al.*, 2017; de Vries *et al.*, 2018a; Mutte *et al.*, 2018; Nishiyama *et al.*, 2018). In land plants, phytohormones interact in converging networks of regulatory circuits (Kohli *et al.*, 2013) many of which are part of the plant perceptron that is put to use when dealing with environmental cues (Scheres and van der Putten, 2017).

Auxin is arguably the most famous phytohormone. Polar transport of auxin is a cornerstone in the development of land plants (e.g. Friml *et al.*, 2003) and it orchestrates developmental processes throughout the plant body (Weijers and Wagner, 2016). In recent years, auxin has further been recognized to be a major player in the adjustment of the cell biology of plants to stress cues (Naser and Shani, 2016; Blakeslee *et al.*, 2019). Auxin has been detected in streptophyte algae and its polar transport has been described in Klebsormidiophyceae and Charophyceae—suggesting that polar auxin transport was an early invention (Cooke *et al.*, 2002; Boot *et al.*, 2012; Hori *et al.*, 2014; Ohtaka *et al.*, 2017). It appears however, that polar auxin transport is not present in all streptophyte algal lineages. Indeed, the KfPIN homolog of *Klebsormidium flaccidum* is an

auxin-specific transporter, which in heterologous experiments localizes in a non-polar manner in the plasma membrane; in *K. flaccidum*, *KjPIN* is localized at the peripheral plasma membrane rather than at positions of cell–cell contact (Skokan *et al.*, 2019). In contrast, in *Chara vulgaris*, auxin transporters appear to be localized in a polar manner (Žabka *et al.*, 2016), which Skokan *et al.* (2019) hypothesized to have evolved convergently in the algae and embryophytes. Further, the canonical auxin perception and transduction pathway probably first emerged in land plants, as inferred from transcriptional responses and the presence of the required signaling components (Mutte *et al.*, 2018).

Streptophyte algae are predicted to have pathways for utilizing phytohormones that are predominantly known as relevant for the response to environmental cues. Foremost among those is the homologous genetic framework for the signaling cascade that all land plants use in ABA-mediated responses (Umezawa *et al.*, 2010; Eklund *et al.*, 2018). ABA is known as a major stress phytohormone; the signaling ABA triggers is involved in responses to abiotic cues such as salt, drought, and temperature (reviewed by Ingram and Bartels, 1996; Shinozaki *et al.*, 2003; Yoshida *et al.*, 2014; Fahad *et al.*, 2015). This ABA signaling cascade consists of a three-component core signaling module (see Cutler *et al.*, 2010) that is a chain of negative regulation: when ABA is present, it binds to a receptor of the PYRABACTIN RESISTANCE1/PYR1-LIKE/REGULATORY COMPONENTS OF ABA RECEPTOR (PYR/PYL/RCAR) family that inhibits the PROTEIN PHOSPHATASE 2C (PP2C) proteins that usually would prevent activity of the SUCROSE NONFERMENTING 1-RELATED PROTEIN KINASEs (SnRKs). The SnRKs are the components that activate the downstream targets, such as ion channels or TFs (Furihata *et al.*, 2006; Geiger *et al.*, 2009). The presence and functional conservation of the interaction of PP2Cs, SnRKs, and downstream targets such as ion channels have been investigated through experimental work for proteins of the streptophyte alga *Klebsormidium* (Holzinger and Becker, 2015; Lind *et al.*, 2015; Shinozawa *et al.*, 2019).

Recently, a transcript probably coding for an orthologous protein of the land plant PYL was detected in *Zygnema circumcarinatum* (de Vries *et al.*, 2018a). The presence of PYR/PYL/RCAR homologs in Zygnematophyceae was recently corroborated by the publication of the first two genomes of Zygnematophyceae (Cheng *et al.*, 2019). Cheng *et al.* (2019) found PYR/PYL/RCAR homologs in one (*Mesotanium endlicherianum*) of the two zygnematophyceae genomes. Interestingly, in the second zygnematophyceae genome the authors analyzed, that of the newly described alga *Spirogloea muscicola* (Cheng *et al.*, 2019), no PYR/PYL/RCAR homolog was found; the same applied to an independent genome study of another Zygnematophyceae, *Penium margaritaceum* (Jiao *et al.*, 2019, Preprint). Hence, the genes of the PYR/PYL/RCAR family were probably gained at the base of the monophyletic group of Zygnematophyceae and land plants (see also the excellent discussion in Cuming, 2019). Functional studies of the protein encoded by the homologous PYL gene found in *Z. circumcarinatum* have shown that ZcPYL does interact with the downstream PP2Cs—but that it does so in an

ABA-independent manner (Sun *et al.*, 2019). The exact regulatory function of PYL:PP2C:SnRK is thus an open question (see also discussions in the last section of this manuscript). It is, however, clear that the presence of these genes has offered fertile ground for the evolution of the canonical ABA signaling cascade that we know from land plants.

Carrying out the essential work downstream of the signal transduction cascades: streptophytic stress responses

Chemodiversity in secondary metabolites is a key trait of land plants. It has been speculated that such a breadth of secondary metabolites has been critical for the success of embryophytes in the challenging environment of *terra firma* (Weng, 2014). A prime source from which a plethora of embryophytic secondary metabolites emerges is the phenylpropanoid pathway, which is the backbone for a range of compounds that can be associated with any abiotic (terrestrial) stressor imaginable (Dixon and Paiva, 1995; Vogt, 2010). All land plants—bryophytes and tracheophytes—use the phenylpropanoid pathway when challenged with stressors (Wolf *et al.*, 2010; Oliva *et al.*, 2015; Albert *et al.*, 2018; Clayton *et al.*, 2018; Carella *et al.*, 2019). Indeed, the phenylpropanoid pathway was thought to have emerged at the base of the clade of embryophytes, thus being an early adaption of embryophytes (Emiliani *et al.*, 2009; Weng, 2014). However, a homologous genetic framework for many components of the core phenylpropanoid pathway was found in many streptophyte algal species when several transcriptomic data sets and one genome were cumulatively investigated (de Vries *et al.*, 2017). This suggests that streptophyte algae could in general be able to synthesize phenylpropanoids—and possibly downstream derivatives, which have been previously reported as a land plant invention (de Vries *et al.*, 2017). Indeed, biochemical and histochemical data accumulated over the years indicate that lignin-like components are present in the cell walls of several streptophyte algae—most prominently in the genus *Coleochaete* (Delwiche *et al.*, 1989; Sørensen *et al.*, 2011). In land plants, the production of lignins can be considered a route that is distal to—but hinging on—the core part of the phenylpropanoid pathway (Vanholme *et al.*, 2012). Together with the genes found in many streptophyte algae, this suggests that the capacity to use core and peripheral routes of the phenylpropanoid pathway arose before phragmoplastophytes emerged. Interestingly, Jiao *et al.* (2019, Preprint) detected flavonoids in the Zygnematophyceae *Penium margaritaceum* despite the absence of homologous genes that might code for the two required upstream enzymes of the core phenylpropanoid pathway in its genome. Altogether, it is likely that the phenylpropanoid pathway was already present—in some fashion—in the earliest land plants. There, it might have acted in the production of compounds that warded off terrestrial stressors. Further, this provided fertile genetic ground for likely gene duplication-based diversification of the pathway routes and the chemodiversity it yields (see Niklas *et al.*, 2017 for an erudite explanation).

High irradiances are a prime challenge in the terrestrial environments. What do extant streptophyte algae tell us about how the earliest land plants might have overcome these? UV irradiances could have been shielded by phenolic compounds produced by the phenylpropanoid pathway (Popper *et al.*, 2011; de Vries *et al.*, 2017), including the recently detected flavonoids (Jiao *et al.*, 2019, Preprint). It is hence conceivable that phenylpropanoid-derived metabolites were part of the UV screen of the earliest land plants.

The green lineage appears to have a nuclear-encoded plastid proteome that is particularly moldable by stressors (Knopp *et al.*, 2019, Preprint). In land plants, a tight communication between the plastid and nucleus adjusts the plastid proteome under photophysiological challenging irradiances via operational retrograde signaling (see also Chan *et al.*, 2016). The evolution of such a tight framework of communication was probably a major benefactor of plant terrestrialization (de Vries *et al.*, 2016). In fact, genes for a land plant-like retrograde signaling—including GENOMES UNCOUPLED 1 (GUN1) and SAL1—from the plastid to the nucleus seem to have emerged at the base of phragmoplastophytes (de Vries *et al.*, 2018a; Nishiyama *et al.*, 2018; Zhao *et al.*, 2019); the earliest land plants probably used these genes in an operational retrograde signaling network for attuning their plastid to terrestrial stressors.

A well-conserved mechanism for protection against high light is the mitigation of excess energy that is generated during photosynthesis through non-photochemical quenching (NPQ). This mechanism involves the dissipation of excess excitation energy of excited singlet chlorophyll as heat, thus preventing ROS formation and reducing the risk of photooxidative damage—especially to the damage-prone PSII (Aro *et al.*, 1993; Müller *et al.*, 2001). NPQ mechanisms exhibit great variability in green algae (Goss and Lepetit, 2015; Christa *et al.*, 2017). Two proteins, namely LHCSR (light-harvesting complex stress-related protein) and PSBS (photosystem II subunit S), play a major role in NPQ. Both proteins help to activate the fastest NPQ component, qE (energy-dependent quenching), by acting as a pH sensor and inducing conformational changes in LHCI complexes (PSBS, Horton *et al.*, 2000; Li *et al.*, 2000, 2004; LHCSR, Peers *et al.*, 2009; Bonente *et al.*, 2011). For some time, it was thought that in green algae qE depended on the action of the LHCSR proteins (Peers *et al.*, 2009), whereas land plants used another protein of the LHC superfamily, PSBS (Li *et al.*, 2000). Yet, this view was changed by two findings: (i) the chlorophyte *Chlamydomonas reinhardtii* PSBS protein was found to be high light induced—an induction that correlates with NPQ (Correa-Galvis *et al.*, 2016); and (ii) mosses also have an NPQ that depends on the action of LHCSRs and PSBS (Alboresi *et al.*, 2010; Gerotto *et al.*, 2012). This means that the differences in PSBS detection along the trajectory of streptophyte algal evolution (see Gerotto and Morosinotto, 2013) must be seen in a new light—possibly reflecting the notorious difficulties in detecting algal PSBS. Nonetheless, this means that the conclusion of Alboresi *et al.* (2010) holds: the earliest land plants probably had a photoprotection mechanisms that included the action of both LHCSRs and PSBS.

Streptophyte algae that live in terrestrial habitats have remarkable photoprotection capacities. For example, terrestrial *Klebsormidium* can tolerate astoundingly high intensities of light without suffering photoinhibition (Karsten *et al.*, 2016; Pierangelini *et al.*, 2017). Yet, these ecophysiological traits differ when considering multiple representatives of Klebsormidiophyceae (Herburger *et al.*, 2016). Hence, it is likely that such photoprotection mechanisms are particularly moldable by environmental factors, hampering inferences over long evolutionary timescales.

Next to high light, desiccation is a major stressor for streptophyte algae that live in terrestrial habitats (Pierangelini *et al.*, 2019). Indeed, tolerance to both often goes hand in hand. Upon water deficiency, carbon fixation is limited but electron flow continues, leading to possible triplet chlorophyll ($^3\text{Chl}^*$) formation which in turn contributes to the formation of the ROS singlet oxygen ($^1\text{O}_2^*$) (Frankel, 1984; Müller *et al.*, 2001). Also, under limited carbon fixation, the biosynthesis of photoprotective molecules is impaired, leading to even more ROS formation (Takahashi and Murata, 2008). Terrestrial streptophyte algae such as *Klebsormidium* and *Zygnema* have a remarkable desiccation tolerance (e.g. Pierangelini *et al.*, 2017; Rippin *et al.*, 2017; Herburger *et al.*, 2019). Their tolerance is based on a range of mechanisms such as protective substance production, cell wall remodeling, formation of specialized cells, and many more (for an excellent overview, see Holzinger and Pichrtová, 2016; for more on cell walls and plant terrestrialization, see also Harholt *et al.*, 2016). Even streptophyte algae that are not considered terrestrial, such as *Coleochaete*, can tolerate desiccation when challenged with it in the lab (Graham *et al.*, 2012). Finding the ability to overcome terrestrial stressors in extant streptophyte algae has important implications: it is likely that streptophyte algae have equally found their way onto land multiple times independently (see also Figs 1, 2). Among terrestrial streptophyte algae, there is a mix of completely independent as well as convergent solutions (many of which we illustrate in the next section) to dealing with terrestrial stressors. An example for divergent solutions is that terrestrial *Klebsormidium* deposits callose under desiccation stress while *Zygnema* does not (Herburger and Holzinger, 2015). However, these differences and independent solutions also highlight that streptophyte algae have a moldable genetic framework for molecular stress responses. Molding this framework eventually formed the network that was also put to use by the earliest land plants.

During the course of evolution, various photosynthetic eukaryotes have found their way onto land. This includes various members of the green lineage that made the wet to dry transition multiple times independently (Lewis and McCourt, 2004). Yet, for example, some diatoms and even cyanobacteria are also terrestrial: photosynthetic organisms that are as distantly related to land plants as can be. These convergently terrestrial organisms must also have found solutions to the challenges of UV, desiccation, high light, etc. In the following we will outline a few of their strategies.

Is the physiology of streptophytes unique?

Photosynthetic organisms from various lineages have settled on land. Only one became the ancestor of the land plants. In an evolutionary context, it is assumed that there were photosynthesizing eukaryotes and prokaryotes on land long before the origin of embryophytes (Raven and Edwards, 2014). How did they overcome the stressors discussed above in the context of plant terrestrialization? Here, again, the diversity of extant photosynthesizing life on land holds illuminating clues.

Desiccation stress responses in green algae can be divided into different categories depending on their function within the organism (Holzinger and Karsten, 2013). One way of responding to desiccation stress is to avoid it. A prime example for desiccation stress avoidance in extreme environmental conditions are desert biological soil crusts (BSCs). Formation of BSCs is a phenomenon that is well known from chlorophytic and streptophytic green algae (e.g. Belnap and Lange, 2001; Flechtner, 2007; Holzinger and Karsten, 2013). Cyanobacteria form BSCs, too. These BSCs are a mixture of sand particles and polysaccharides which are excreted by the cyanobacteria and are able to withstand extreme environmental desiccation, temperature changes, and very high irradiance (Mager and Thomas, 2011; Ferrenberg *et al.*, 2015). Particularly noteworthy is the cyanobacterium *Leptolyngbya ohadii* that is native to deserts (Raanan *et al.*, 2016; Oren *et al.*, 2019).

Leptolyngbya ohadii employs a range of physiological mechanisms to go through cycles of dehydration and recovery during rewetting: before dehydration, phytochromes and cryptochromes act as signal transmitters, sensing dawn illumination and thus preparing the cell for impending loss of water (Oren *et al.*, 2017). During dehydration, trehalose-based osmotic adjustment and anhydrobiotic protection appear to be key (Murik *et al.*, 2017)—although recent data suggest that their role might be limited (Oren *et al.*, 2019). The dehydration/rewetting cycles induce pronounced transcriptional reprogramming. Genes of *L. ohadii* involved in photoprotection—such as OCP (Leverenz *et al.*, 2015)—were up-regulated during dehydration while genes involved in biosynthesis of photosynthetic components were down-regulated (Oren *et al.*, 2019). Cellular activity (photosynthesis) in *L. ohadii* and other cyanobacteria was measured a short time after rewetting: the results suggested a complex gene regulatory network that is highly adapted to dehydration/rewetting cycles being reflected in swift modulation of the photosynthetic machinery (Bar-Eyal *et al.*, 2015; Oren *et al.*, 2019). It was proposed that intrinsically disordered proteins (IDPs) play a possible role in protecting and stabilizing RNA during desiccation (Oren *et al.*, 2019). IDPs have been studied regarding their role in desiccation tolerance of tardigrades (*Milnesium tardigradum*; Boothby *et al.*, 2017) and might have a similar role in Chloroplastida, too (Sun *et al.*, 2013; Y. Zhang *et al.*, 2018; see also Niklas *et al.*, 2018). Hence, even in terrestrial cyanobacteria—which reside in another domain of life—mechanisms that function similarly in terrestrial chlorophytes and streptophytes can be found.

Protective responses to desiccation stress include formation of a modified cell wall. Some Trebouxiophyceae do so palpably. *Prasiola crista* subsp. *antarctica* grows, as the name implies, in the

supralittoral of Antarctica and possesses a very thick cell wall that protects the alga from osmotic stress during desiccation (Jacob *et al.*, 1992). The cell wall in *Prasiola* contains a large amount of pectin (Jacob *et al.*, 1992). Pectins are important components of the cell wall of land plants and streptophyte algae (Ridley *et al.*, 2001; Sørensen *et al.*, 2011) but have also been shown to be involved in desiccation stress responses in the ulvophyte *Ulva compressa* by providing a flexible cell wall (Holzinger *et al.*, 2015). Under hypoosmotic conditions, the flexible nature of the cell wall prevents the algal cells from swelling excessively, while under hyperosmotic conditions it prevents plasmolysis—a feature that is beneficial not only under desiccation stress but also under salinity stress which is a frequent stressor considering the habitat of *P. crista* subsp. *antarctica* (Jacob *et al.*, 1992). Hence, both chlorophytes and streptophytes have evolved mechanisms for cell wall modification under desiccation stress.

The ability of algae to form symbiotic interactions is assumed to have been key in the conquest of land by plants (Delaux *et al.*, 2013; Field *et al.*, 2015). While land plants have unique beneficial symbionts (foremost the arbuscular mycorrhiza), other green organisms are not without friends. For example, many species of Trebouxiophyceae are lichen-forming. Lichens are highly resistant to desiccation and are thus able to stay in a dehydrated state for an extended amount of time until rehydration. In order for a lichen to survive under such extreme conditions, the underlying molecular regulatory network must be extremely fine-tuned and adapted to dehydration/rehydration cycles. For example, dehydration of DNA resulting in strand breaks would mean death for the lichen (Dose *et al.*, 1992). Carniel *et al.* (2016) provide a good example for transcriptomic regulation in the lichen photobiont *Trebouxia gelatinosa*; their data reveal a different gene regulatory set-up in various categories such as cell wall modifications, photosynthetic apparatus, or oxidative stress response during dehydration/rehydration cycles (Carniel *et al.*, 2016). In this context, photoprotective processes are crucial because in desiccated photosynthetic tissues there is a high risk of excessive ROS formation (Smirnoff, 1993; Scheibe and Beck, 2011). Rapid adjustments of the antioxidant homeostasis appear to be a mechanisms that lichens employ under desiccation stress (Calatayud *et al.*, 1997; Zorn *et al.*, 2001; Kranner *et al.*, 2003); such antioxidants include zeaxanthin, which can deactivate excited chlorophylls (Jahns and Holzwarth, 2012)—thus substances that are found conserved across plastid-bearing eukaryotes (Dautermann and Lohr, 2017).

The aforementioned PSBS and LHCSR are LHC-like proteins that are key for the response to high light. An additional important group of LHC-like proteins are the early light-induced proteins (ELIPs), which are swiftly up-regulated in response to high light stress as well as also being up-regulated in response to dehydration stress and by ABA (Bartels *et al.*, 1992; Pötter and Kloppstech, 1993; Zeng *et al.*, 2002; Dinakar and Bartels, 2013). ELIPs are hallmark stress-responsive proteins in a variety of oxygenic photosynthetic organisms from cyanobacteria, to algae, to land plants (Heddad and Adamska, 2002). Hutin *et al.* (2003) demonstrated that ELIPs have an essential photoprotective role by using the *A. thaliana* mutant *chaos*,

which is not able to translocate LHC-type proteins such as ELIP via the chloroplast signal recognition particle (CpSRP) pathway to the thylakoid membranes (see also Hutin *et al.*, 2002). The authors observed leaf bleaching and photooxidative damage in *chaos* mutants challenged with high-light cues. The data of Hutin *et al.* (2003) thus revealed the photoprotective role of ELIPs. It was further suggested that ELIPs have an influence on chlorophyll biosynthesis, thus indirectly preventing chlorophyll accumulation under high light conditions and thereby also preventing ROS formation (Hutin *et al.*, 2003; Tzvetkova–Chevolleau *et al.*, 2007).

The formation of sunscreens is a powerful protection against UV irradiance. Above, we illustrated how the biosynthesis of phenylpropanoid-derived metabolites is involved in this mechanism in streptophytes—and discussed how such metabolites might have been critical for streptophyte terrestrialization. Yet, other terrestrial organisms have sunscreens, too. In some cyanobacteria, scytonemin acts as a very potent UVA sunscreen that protects the cells from near UV and blue radiation (Gao and Garcia-Pichel, 2011). It accumulates as a stable pigment in extracellular polysaccharide sheaths of cyanobacteria and possesses various convenient characteristics. Scytonemin changes dependent on the redox status, stays active even under physiological inactive conditions, and is capable in performing strong absorption in the UVA range due to its ring structure with conjugated double bonds (Garcia-Pichel and Castenholz, 1991; Proteau *et al.*, 1993; Gao and Garcia-Pichel, 2011). The pigment is typically found in biological soil crusts or epilithic biofilms (Gao and Garcia-Pichel, 2011). While many questions regarding its biosynthesis and regulation remain, the exposure of cyanobacterial cells to UVA irradiance, however, appears to be correlated with scytonemin production (Garcia-Pichel and Castenholz, 1991; see also Ehling-Schulz *et al.*, 1997).

In chlorophyte algae, mycosporine-like amino acids (MAAs) act as sunscreens when exposed to solar UV radiation. For example, an accumulation of MAAs was found to be advantageous against UV irradiance in the aeroterrestrial green algae *Stichococcus* sp. and *Chlorella luteoviridis* when compared with two green algae from soil with a different MAA set-up (Karsten *et al.*, 2007). Such aeroterrestrial species possess a unique type of MAA (324 nm MAA) which is only found in Trebouxiophyceae—including the alga *P. crista* spp. *antarctica* (Hoyer *et al.*, 2001; Karsten *et al.*, 2005). In the streptophyte algal class of Klebsormidiophyceae, we also find the production of 324 nm MAA—yet, despite their similar absorption spectrum, these are different substances (Kitzing and Karsten, 2015). This provides yet another intriguing example of convergent evolution.

So, what makes streptophytes special? There are multiple lineages of algae that made their way onto land by mastering its stressors. Post-hoc, we see that only the progenitors of the last common ancestor of embryophytes gave rise to a lineage that globally conquered land. Why do only they dominate the terrestrial flora? A specific (embryophyte-like) physiology that aided them in dealing with stressors will not be the only reason. However, it has to be part of the answer. Dealing with terrestrial stressors was not only under selection in the earliest land plants but is under selection still. Like the core processes in

energy metabolism (such as the citrate cycle) that are essential, core response mechanisms to stressors are continually tested by adverse terrestrial conditions and are essential for survival in the terrestrial habitat. The question is which building blocks of that core were present in the earliest land plants.

Rewiring of an ancient molecular physiology was a likely facilitator of the success of land plants

The responses of land plants to stressors hinges on complex regulatory networks. In such stress-relevant networks, nodes are genes/proteins that are connected by edges that circumscribe an interaction—such as phosphorylation and transcriptional activation—that are triggered and/or modulated by environmental stimuli. Dissecting ancient pathways and setting them into context with land plant data enables us to trace the evolution of such networks across the streptophyte tree of life. Correlative data, such as genomes, have indicated the presence of many genes (the nodes) that could act in stress-relevant regulatory networks—but what about the wires that connect them (the edges)? To understand these, more involved functional analyses are required.

ABA is a key modulator of stress response. Above, we outlined that streptophyte algae have most—and Zygnematiophyceae all—homologous genes of the canonical cascade that land plants use for ABA signaling (deVries *et al.*, 2018a; Cheng *et al.*, 2019; Sun *et al.*, 2019). What does this mean for the evolution of the laying of the wires that underpin ABA signaling? Although *Klebsormidium* completely lacks a gene for the ABA receptor, Holzinger *et al.* (2014) found the other components (PP2C, SnRK2s, and AREBs) to be responsive to drought (see also Holzinger and Becker, 2015). Furthermore, the genes of *Klebsormidium* that are homologous to ABA signaling components complement the respective mutants of *A. thaliana* and *P. patens* in heterologous experiments (Lind *et al.*, 2015; Shinozawa *et al.*, 2019); for example, expressing *Klebsormidium nitens* SnRK2.6 in *A. thaliana* protoplasts that are deficient in SnRK2.2, 2.3, and 2.6 (*snrk2.2/2.3/2.6* triple mutants) rescues the transduction of ABA-induced gene expression (Lind *et al.*, 2015). Hence, the specific wires of the interaction required for the ABA signaling cascade that are downstream of the ABA–PYL interaction appear to be conserved between Klebsormidiophyceae and land plants. This means that the cascade is probably hundreds of millions of years older than its bona fide ABA dependency mediated by PYLs. Recently, another piece was added to the puzzle of the evolution of the ABA signaling cascade: through *in vitro* and heterologous work, Sun *et al.* (2019) showed that the PYL homolog found in *Z. circumcarinatum* does regulate its downstream target—the PP2Cs. Yet, their data also show that this regulation happens in an ABA-independent manner. What this means is that the bona fide ABA signaling cascade, which consists of PYLs:PP2Cs:SnRKs, has evolved in a modular fashion—with components being successively plugged in. The ABA dependency of the cascade—another addition—has probably evolved along the trajectory from the earliest land plants to the last

common ancestor of land plants (see also the discussion on this topic in Sun *et al.*, 2019). Whether there is nonetheless an ABA dependency of parts of the signaling cascade in algae mediated through other means—such as the *ABA NON-RESPONSIVE* kinases that are only present in non-flowering plants (Stevenson *et al.*, 2016)—remains to be investigated. However, it offers an explanation for how conserved stress tolerance-conferring mechanisms such as LEA accumulation can be part of the same cascade that we now think of as ABA mediated: they might have been part of the downstream response triggered by the same regulatory cascade conserved for hundreds of millions of years, in which they have been functional and riding along since before the signal transduction chain was under the control of ABA.

The signaling network that mediates the action of the growth phytohormone auxin has a similar story to tell. A series of recent studies [Flores-Sandoval *et al.* (2018); Mutte *et al.* (2018); Martin-Arevalillo *et al.* (2019)] proposed an evolutionary history that features a modular build-up for the nuclear auxin response pathway that is mediated by AUXIN RESPONSE FACTORS (ARFs), the TRANSPORT INHIBITOR RESPONSE 1/AUXIN SIGNALING F-BOX (TIR/AFB), AUXIN/INDOLE-3-ACETIC ACID PROTEIN (AUX/IAA) system: class C ARFs are present in streptophyte algae—and perhaps were even present in the last common ancestor of streptophytes (see also Wang *et al.*, 2019); the common ancestor of Coleochaetophyceae, Zygnematophyceae, and land plants gained the single co-ortholog (A/B ARF) of the classes of A and B ARFs known from land plants. In the absence of TIR/AFB and AUX/IAA, these algal ARFs act in an auxin-independent manner (Martin-Arevalillo *et al.*, 2019; see also Mutte *et al.*, 2018). In land plants, the A/B ARF diverged into the class A and class B ARFs (Flores-Sandoval *et al.*, 2018; Mutte *et al.*, 2018; Martin-Arevalillo *et al.*, 2019). Onto this diversified system, the TIR/AFB and AUX/IAA were plugged, rendering the whole system auxin dependent (Flores-Sandoval *et al.*, 2018; Mutte *et al.*, 2018; Martin-Arevalillo *et al.*, 2019). Thus, adding one regulatory mechanism upstream was sufficient to turn a conserved gene regulatory cascade into a phytohormone-dependent cascade. Similar to the aforementioned case of ABA-mediated signaling, gaining such phytohormone dependency probably occurred along the trajectory between the earliest land plants and the last common ancestor of land plants.

Recent functional studies on the liverwort model system *M. polymorpha* have illustrated how a different type of modular evolution of a phytohormone pathway can occur—in this case not concerning a protein that is plugged in, but a different effector molecule. Of course, such changes ultimately hinge on differences on the protein level, too; that is, in the binding pocket for the input molecule. Monte *et al.* (2018) investigated the origin of jasmonic acid (JA) perception by focusing on the COI1 receptor. Instead of sensing JA-Ile, the COI1 protein of *M. polymorpha* senses dinor-12-oxo-phytodienoic acid (dinor-OPDA), which emerges from an earlier branching point in the JA biosynthesis pathway (Monte *et al.*, 2018). Hence, while the entire pathway for the canonical, JA-Ile-based, perception of JA is present in the bryophyte *M. polymorpha*, it does not work

in the same way as we know it from angiosperms. Plugging in a different molecule at the uppermost layer of the signaling cascade hence has the ability to change the entire co-evolutionary relationship between phytohormone biosynthesis and its perception. Even more extreme cases of rewiring can be expected to have taken place in streptophyte algae, which are hundreds of millions of years divergent from land plants. Thus, the gain of a specific phytohormone dependency can also entail shifts in input signal. Illuminating whether such shifts in input signal underpinned the evolution of other phytohormone signaling cascades (such as ABA) is an exciting avenue for future research.

Rewiring is not limited to signaling pathways. Secondary metabolites of land plants are known for their chemodiverse and lineage-specific—sometimes species-specific—secondary metabolic fingerprints (see, for example, Dudareva *et al.*, 2004). During the course of evolution, the same building block-producing backbone pathways have been rewired to give rise to tens of thousands of different secondary metabolites. For example, Berland *et al.* (2019) recently showed that the liverwort model system *M. polymorpha* produces a novel class of anthocyanins, which they termed auronidins. Auronidins appear to be derived from anthocyanin but emerge from a novel route via auronones (Berland *et al.*, 2019). This highlights the complexities one has to tackle when inferring the secondary metabolism of the earliest land plants. Since the genes for the phenylpropanoid biosynthesis pathway are present in streptophyte algae (de Vries *et al.*, 2017) and building blocks have been detected even in chlorophytes (Goiris *et al.*, 2014), we can expect that the earliest land plants used a plethora of secondary metabolites derived from these routes. Inferring the routing of the pathways, however, requires more than the mere knowledge of the presence of some genes or some metabolites. A functional dissection of the biosynthetic pathways across the streptophyte tree of life—especially in the streptophyte algae—is needed.

Conclusion

The lineage of embryophytes has conquered land. Only through a fortuitous combination of traits did they succeed in this conquest. These traits include complex networks for stress response that are, in their elaboration, probably limited to land plants. Yet, often, the integral components and/or decisive nodes and wires of these networks were already present—they represent complete and functioning building blocks ready for co-option. Moreover, as of yet, we do not know how most of these components function in their own environment; that is, how are they wired in streptophyte algae? It may very well be that the present components have the same functions as in land plants, yet their interactors may be different in streptophyte algae (e.g. different receptor specificity or different downstream components). Likewise, a completely different function for the streptophyte algal components underlying a network is also entirely possible. One thing, however, is very clear; these conserved proteins that participate in the stress signaling networks in land plants will not be without a function in streptophyte algae. Comparative studies that dissect the routing of upstream (e.g. signaling) and downstream (e.g.

biosynthesis of stress protectants) stress response pathways across streptophytes will illuminate how their co-options were realized in the earliest land plants.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 852725), supporting work in the lab of JdV through the ERC Starting Grant 'TerreStriAL'. We thank Debbie Maizels (Zoobotanica Scientific Illustration: www.scientific-art.com) for her marvelous work on Figs 1 and 2.

References

- Adl SM, Bass D, Lane CE, *et al.* 2019. Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* **66**, 4–119.
- Adl SM, Simpson AG, Farmer MA, *et al.* 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology* **52**, 399–451.
- Albert NW, Thrimawithana AH, McGhie TK, Clayton WA, Deroles SC, Schwinn KE, Bowman JL, Jordan BR, Davies KM. 2018. Genetic analysis of the liverwort *Marchantia polymorpha* reveals that R2R3MYB activation of flavonoid production in response to abiotic stress is an ancient character in land plants. *New Phytologist* **218**, 554–566.
- Alboresi A, Gerotto C, Giacometti GM, Bassi R, Morosinotto T. 2010. *Physcomitrella patens* mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization. *Proceedings of the National Academy of Sciences, USA* **107**, 11128–11133.
- Archibald JM. 2015. Genomic perspectives on the birth and spread of plastids. *Proceedings of the National Academy of Sciences, USA* **112**, 10147–10153.
- Aro EM, Virgin I, Andersson B. 1993. Photoinhibition of photosystem II. Inactivation, protein damage and turnover. *Biochimica et Biophysica Acta* **1143**, 113–134.
- Arrighi JF, Barre A, Ben Amor B, *et al.* 2006. The *Medicago truncatula* lysin [corrected] motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. *Plant Physiology* **142**, 265–279.
- Bar-Eyal L, Eisenberg I, Faust A, *et al.* 2015. An easily reversible structural change underlies mechanisms enabling desert crust cyanobacteria to survive desiccation. *Biochimica et Biophysica Acta* **1847**, 1267–1273.
- Bartels D, Hanke C, Schneider K, Michel D, Salamini F. 1992. A desiccation-related Elip-like gene from the resurrection plant *Craterostigma plantagineum* is regulated by light and ABA. *The EMBO Journal* **11**, 2771–2778.
- Becker B, Marin B. 2009. Streptophyte algae and the origin of embryophytes. *Annals of Botany* **103**, 999–1004.
- Belnap J, Lange OL. 2001. *Biological soil crusts: structure, function and management*. 7 Berlin Heidelberg: Springer-Verlag.
- Bengtson S, Sallstedt T, Belivanova V, Whitehouse M. 2017. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biology* **15**, e2000735.
- Berbee ML, James TY, Strullu-Derrien C. 2017. Early diverging fungi: diversity and impact at the dawn of terrestrial life. *Annual Review of Microbiology* **71**, 41–60.
- Berland H, Albert NW, Stavland A, *et al.* 2019. Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants. *Proceedings of the National Academy of Sciences, USA* **116**, 20232–20239.
- Blakeslee JJ, Spatola Rossi T, Kriechbaumer V. 2019. Auxin biosynthesis: spatial regulation and adaptation to stress. *Journal of Experimental Botany* **70**, 5041–5049.
- Bonente G, Ballottari M, Truong TB, Morosinotto T, Ahn TK, Fleming GR, Niyogi KK, Bassi R. 2011. Analysis of LhcSR3, a protein essential for feedback de-excitation in the green alga *Chlamydomonas reinhardtii*. *PLoS Biology* **9**, e1000577.
- Boot KJ, Libbenga KR, Hille SC, Offringa R, van Duijn B. 2012. Polar auxin transport: an early invention. *Journal of Experimental Botany* **63**, 4213–4218.
- Boothby TC, Tapia H, Brozena AH, Piszkievicz S, Smith AE, Giovannini I, Rebecchi L, Pielak GJ, Koshland D, Goldstein B. 2017. Tardigrades use intrinsically disordered proteins to survive desiccation. *Molecular Cell* **65**, 975–984.e5.
- Bowman JL, Sakakibara K, Furumizu C, Dierschke T. 2016. Evolution in the cycles of life. *Annual Review of Genetics* **50**, 133–154.
- Bressendorff S, Azevedo R, Kenchappa CS, Ponce de León I, Olsen JV, Rasmussen MW, Erbs G, Newman MA, Petersen M, Mundy J. 2016. An innate immunity pathway in the moss *Physcomitrella patens*. *The Plant Cell* **28**, 1328–1342.
- Breuninger H, Thamm A, Streubel S, Sakayama H, Nishiyama T, Dolan L. 2016. Diversification of a transcription factor family led to the evolution of antagonistically acting genetic regulators of root hair growth. *Current Biology* **26**, 1622–1628.
- Butterfield NJ. 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404.
- Calatayud A, Deltoro VI, Barreno E, Del Valle-Tascon S. 1997. Changes in *in vivo* chlorophyll fluorescence quenching in lichen thalli as a function of water content and suggestion of zeaxanthin-associated photoprotection. *Physiologia Plantarum* **101**, 93–102.
- Carella P, Gogleva A, Hoey DJ, Bridgen AJ, Stolze SC, Nakagami H, Schornack S. 2019. Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Current Biology* **29**, 2282–2294.e5.
- Carniel FC, Gerdol M, Montagner A, Banchi E, De Moro G, Manfrin C, Muggia L, Pallavicini A, Tretiac M. 2016. New features of desiccation tolerance in the lichen photobiont *Trebouxia gelatinosa* are revealed by a transcriptomic approach. *Plant Molecular Biology* **91**, 319–339.
- Catarino B, Hetherington AJ, Emms DM, Kelly S, Dolan L. 2016. The stepwise increase in the number of transcription factor families in the pre-cambrian predated the diversification of plants on land. *Molecular Biology and Evolution* **33**, 2815–2819.
- Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ. 2016. Learning the language of the chloroplast: retrograde signaling and beyond. *Annual Review of Plant Biology* **67**, 25–53.
- Chater CCC, Caine RS, Fleming AJ, Gray JE. 2017. Origins and evolution of stomatal development. *Plant Physiology* **174**, 624–638.
- Cheng S, Xian W, Fu Y, *et al.* 2019. Genomes of subaerial zygomatophyceae provide insights into land plant evolution. *Cell* **179**, 1057–1067.e14.
- Christa G, Cruz S, Jahns P, de Vries J, Cartaxana P, Esteves AC, Serôdio J, Gould SB. 2017. Photoprotection in a monophyletic branch of chlorophyte algae is independent of energy-dependent quenching (qE). *New Phytologist* **214**, 1132–1144.
- Clayton WA, Albert NW, Thrimawithana AH, *et al.* 2018. UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *The Plant Journal* **96**, 503–517.
- Cooke TJ, Poli D, Szein AE, Cohen JD. 2002. Evolutionary patterns in auxin action. *Plant Molecular Biology* **49**, 319–338.
- Correa-Galvis V, Redekop P, Guan K, Griess A, Truong TB, Wakao S, Niyogi KK, Jahns P. 2016. Photosystem II subunit PsbS is involved in the induction of LHCSR protein-dependent energy dissipation in *Chlamydomonas reinhardtii*. *Journal of Biological Chemistry* **291**, 17478–17487.
- Cuming AC. 2019. The evolution of ABA signalling. *Advances in Botanical Research* **92**, 281–313.
- Cuming AC, Cho SH, Kamisugi Y, Graham H, Quatrano RS. 2007. Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*. *New Phytologist* **176**, 275–287.
- Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. 2010. Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology* **61**, 651–679.
- Dautermann O, Lohr M. 2017. A functional zeaxanthin epoxidase from red algae shedding light on the evolution of light-harvesting carotenoids and the xanthophyll cycle in photosynthetic eukaryotes. *The Plant Journal* **92**, 879–891.

- Delaux PM, Hetherington AJ, Coudert Y, et al.** 2019. Reconstructing trait evolution in plant evo–devo studies. *Current Biology* **29**, R1105–R1121.
- Delaux PM, Radhakrishnan GV, Jayaraman D, et al.** 2015. Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences, USA* **112**, 13390–13395.
- Delaux PM, Séjalon-Delmas N, Bécard G, Ané JM.** 2013. Evolution of the plant–microbe symbiotic ‘toolkit’. *Trends in Plant Science* **18**, 298–304.
- Delaux PM, Varala K, Edger PP, Coruzzi GM, Pires JC, Ané JM.** 2014. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PLoS Genetics* **10**, e1004487.
- Delaux PM, Xie X, Timme RE, Puech-Pages V, Dunand C, Lecompte E, Delwiche CF, Yoneyama K, Bécard G, Séjalon-Delmas N.** 2012. Origin of strigolactones in the green lineage. *New Phytologist* **195**, 857–871.
- Delwiche CF, Cooper ED.** 2015. The evolutionary origin of a terrestrial flora. *Current Biology* **25**, R899–R910.
- Delwiche CF, Graham LE, Thomson N.** 1989. Lignin-like compounds and sporopollenin *Coleochaete*, an algal model for land plant ancestry. *Science* **245**, 399–401.
- Delwiche CF, Karol KG, Cimino MT, Sytsma KJ.** 2002. Phylogeny of the genus *Coleochaete* (Coleochaetales, charophyta) and related taxa inferred by analysis of the chloroplast gene *rbcl*. *Journal of Phycology* **38**, 394–403.
- de Vries J, Archibald JM.** 2017. Endosymbiosis: did plastids evolve from a freshwater cyanobacterium? *Current Biology* **27**, R103–R105.
- de Vries J, Archibald JM.** 2018. Plant evolution: landmarks on the path to terrestrial life. *New Phytologist* **217**, 1428–1434.
- de Vries J, Curtis BA, Gould SB, Archibald JM.** 2018a. Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proceedings of the National Academy of Sciences, USA* **115**, E3471–E3480.
- de Vries J, de Vries S, Slamovits CH, Rose LE, Archibald JM.** 2017. How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant & Cell Physiology* **58**, 934–945.
- de Vries J, Gould SB.** 2018. The monoplasic bottleneck in algae and plant evolution. *Journal of Cell Science* **131**, jcs203414.
- de Vries J, Stanton A, Archibald JM, Gould SB.** 2016. Streptophyte terrestrialization in light of plastid evolution. *Trends in Plant Science* **21**, 467–476.
- de Vries S, de Vries J, von Dahlen JK, Gould SB, Archibald JM, Rose LE, Slamovits CH.** 2018b. On plant defense signaling networks and early land plant evolution. *Communicative & Integrative Biology* **11**, 1–14.
- Dinakar C, Bartels D.** 2013. Desiccation tolerance in resurrection plants: new insights from transcriptome, proteome and metabolome analysis. *Frontiers in Plant Science* **4**, 482.
- Dixon RA, Paiva NL.** 1995. Stress-induced phenylpropanoid metabolism. *The Plant Cell* **7**, 1085–1097.
- Dose K, Bieger-Dose A, Labusch M, Gill M.** 1992. Survival in extreme dryness and DNA-single-strand breaks. *Advances in Space Research* **12**, 221–229.
- Duckett JG, Pressel S.** 2018. The evolution of the stomatal apparatus: intercellular spaces and sporophyte water relations in bryophytes—two ignored dimensions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20160498.
- Dudareva N, Pichersky E, Gershenzon J.** 2004. Biochemistry of plant volatiles. *Plant Physiology* **135**, 1893–1902.
- Dupuy L, Mackenzie J, Haseloff J.** 2010. Coordination of plant cell division and expansion in a simple morphogenetic system. *Proceedings of the National Academy of Sciences, USA* **107**, 2711–2716.
- Dure L 3rd, Greenway SC, Galau GA.** 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry* **20**, 4162–4168.
- Ehling-Schulz M, Bilger W, Scherer S.** 1997. UV-B-induced synthesis of photoprotective pigments and extracellular polysaccharides in the terrestrial cyanobacterium *Nostoc commune*. *Journal of Bacteriology* **179**, 1940–1945.
- Eklund DM, Kanei M, Flores-Sandoval E, Ishizaki K, Nishihama R, Kohchi T, Lagercrantz U, Bhalerao RP, Sakata Y, Bowman JL.** 2018. An evolutionarily conserved abscisic acid signaling pathway regulates dormancy in the liverwort *Marchantia polymorpha*. *Current Biology* **28**, 3691–3699.
- Eme L, Sharpe SC, Brown MW, Roger AJ.** 2014. On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *Cold Spring Harbor Perspectives in Biology* **6**, a016139.
- Emiliani G, Fondi M, Fani R, Gribaldo S.** 2009. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biology Direct* **4**, 7.
- Fahad S, Hussain S, Matloob A, et al.** 2015. Phytohormones and plant responses to salinity stress: a review. *Plant Growth Regulation* **75**, 391–404.
- Ferrenberg S, Reed SC, Belnap J.** 2015. Climate change and physical disturbance cause similar community shifts in biological soil crusts. *Proceedings of the National Academy of Sciences, USA* **112**, 12116–12121.
- Field KJ, Pressel S.** 2018. Unity in diversity: structural and functional insights into the ancient partnerships between plants and fungi. *New Phytologist* **220**, 996–1011.
- Field KJ, Pressel S, Duckett JG, Rimington WR, Bidartondo MI.** 2015. Symbiotic options for the conquest of land. *Trends in Ecology & Evolution* **30**, 477–486.
- Flechtner VR.** 2007. North American desert microbiotic soil crust communities: diversity despite challenge. In: Seckbach J, ed. *Algae and cyanobacteria in extreme environments*. Berlin: Springer, 537–551.
- Flores-Sandoval E, Eklund DM, Hong SF, et al.** 2018. Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in *Marchantia polymorpha*. *New Phytologist* **218**, 1612–1630.
- Frankel EN.** 1984. Chemistry of free radical and singlet oxidation of lipids. *Progress in Lipid Research* **23**, 197–221.
- Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jürgens G.** 2003. Efflux-dependent auxin gradients establish the apical–basal axis of Arabidopsis. *Nature* **426**, 147–153.
- Furihata T, Maruyama K, Fujita Y, Umezawa T, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K.** 2006. Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proceedings of the National Academy of Sciences, USA* **103**, 1988–1993.
- Gao Q, Garcia-Pichel F.** 2011. Microbial ultraviolet sunscreens. *Nature Reviews: Microbiology* **9**, 791–802.
- Gao Y, Wang W, Zhang T, Gong Z, Zhao H, Han GZ.** 2018. Out of water: the origin and early diversification of plant R-genes. *Plant Physiology* **177**, 82–89.
- Garcia-Pichel F, Castenholz RW.** 1991. Characterization and biological implications of scytonemin, a cyanobacterial sheath pigment. *Journal of Phycology* **27**, 395–409.
- Geiger D, Scherzer S, Mumm P, et al.** 2009. Activity of guard cell anion channel SLAC1 is controlled by drought-stress signaling kinase–phosphatase pair. *Proceedings of the National Academy of Sciences, USA* **106**, 21425–21430.
- Gerotto C, Alboresi A, Giacometti GM, Bassi R, Morosinotto T.** 2012. Coexistence of plant and algal energy dissipation mechanisms in the moss *Physcomitrella patens*. *New Phytologist* **196**, 763–773.
- Gerotto C, Morosinotto T.** 2013. Evolution of photoprotection mechanisms upon land colonization: evidence of PSBS-dependent NPQ in late Streptophyte algae. *Physiologia Plantarum* **149**, 583–598.
- Goiris K, Muylaert K, Voorspoels S, Noten B, De Paepe D, E Baart GJ, De Cooman L.** 2014. Detection of flavonoids in microalgae from different evolutionary lineages. *Journal of Phycology* **50**, 483–492.
- Gómez-Gómez L, Boller T.** 2000. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell* **5**, 1003–1011.
- Goss R, Lepetit B.** 2015. Biodiversity of NPQ. *Journal of Plant Physiology* **172**, 13–32.
- Graham LE, Arancibia-Avila P, Taylor WA, Strother PK, Cook ME.** 2012. Aeroterrestrial *Coleochaete* (Streptophyta, Coleochaetales) models early plant adaptation to land. *American Journal of Botany* **99**, 130–144.
- Han GZ.** 2019. Origin and evolution of the plant immune system. *New Phytologist* **222**, 70–83.
- Harholt J, Moestrup Ø, Ulvskov P.** 2016. Why plants were terrestrial from the beginning. *Trends in Plant Science* **21**, 96–101.
- Harrison CJ.** 2017. Development and genetics in the evolution of land plant body plans. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, 20150490.

- Heddad M, Adamska I.** 2002. The evolution of light stress proteins in photosynthetic organisms. *Comparative and Functional Genomics* **3**, 504–510.
- Herburger K, Holzinger A.** 2015. Localization and quantification of callose in the streptophyte green algae *Zygnema* and *Klebsormidium*: correlation with desiccation tolerance. *Plant & Cell Physiology* **56**, 2259–2270.
- Herburger K, Karsten U, Holzinger A.** 2016. *Entransia* and *Hormidiella*, sister lineages of *Klebsormidium* (Streptophyta), respond differently to light, temperature, and desiccation stress. *Protoplasma* **253**, 1309–1323.
- Herburger K, Xin A, Holzinger A.** 2019. Homogalacturonan accumulation in cell walls of the green alga *Zygnema* sp. (Charophyta) increases desiccation resistance. *Frontiers in Plant Science* **10**, 540.
- Hetherington AJ, Dolan L.** 2017. The evolution of lycopsid rooting structures: conservatism and disparity. *New Phytologist* **215**, 538–544.
- Hirsch S, Oldroyd GE.** 2009. GRAS-domain transcription factors that regulate plant development. *Plant Signaling & Behavior* **4**, 698–700.
- Hoffmann L.** 1989. Algae of terrestrial habitats. *Botanical Review* **55**, 77–105.
- Holdsworth MJ, Bentsink L, Soppe WJ.** 2008. Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytologist* **179**, 33–54.
- Holzinger A, Becker B.** 2015. Desiccation tolerance in the streptophyte green alga *Klebsormidium*: the role of phytohormones. *Communicative & Integrative Biology* **8**, e1059978.
- Holzinger A, Herburger K, Kaplan F, Lewis LA.** 2015. Desiccation tolerance in the chlorophyte green alga *Ulva compressa*: does cell wall architecture contribute to ecological success? *Planta* **242**, 477–492.
- Holzinger A, Kaplan F, Blaas K, Zechmann B, Komsic-Buchmann K, Becker B.** 2014. Transcriptomics of desiccation tolerance in the streptophyte green alga *Klebsormidium* reveal a land plant-like defense reaction. *PLoS One* **9**, e110630.
- Holzinger A, Karsten U.** 2013. Desiccation stress and tolerance in green algae: consequences for ultrastructure, physiological and molecular mechanisms. *Frontiers in Plant Science* **4**, 327.
- Holzinger A, Pichtová M.** 2016. Abiotic stress tolerance of charophyte green algae: new challenges for omics techniques. *Frontiers in Plant Science* **7**, 678.
- Hori K, Maruyama F, Fujisawa T, et al.** 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications* **5**, 3978.
- Horst NA, Katz A, Pereman I, Decker EL, Ohad N, Reski R.** 2016. A single homeobox gene triggers phase transition, embryogenesis and asexual reproduction. *Nature Plants* **2**, 15209.
- Horton P, Ruban AV, Wentworth M.** 2000. Allosteric regulation of the light-harvesting system of photosystem II. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**, 1361–1370.
- Hoyer K, Karsten U, Sawall T, Wiencke C.** 2001. Photoprotective substances in Antarctic macroalgae and their variation with respect to depth distribution, different tissues and developmental stages. *Marine Ecology Progress Series* **211**, 117–129.
- Hutin C, Havaux M, Carde JP, Kloppstech K, Meierhoff K, Hoffman N, Nussaume L.** 2002. Double mutation cpSRP43/cpSRP54 is necessary to abolish the cpSRP pathway required for thylakoid targeting of the light-harvesting chlorophyll proteins. *The Plant Journal* **29**, 531–543.
- Hutin C, Nussaume L, Moise N, Moya I, Kloppstech K, Havaux M.** 2003. Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proceedings of the National Academy of Sciences, USA* **100**, 4921–4926.
- Ikegaya H, Sonobe S, Murakami K, Shimmen T.** 2008. Rhizoid differentiation of *Spirogyra* is regulated by substratum. *Journal of Plant Research* **121**, 571–579.
- Ingram J, Bartels D.** 1996. The molecular basis of dehydration tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**, 377–403.
- Jackson C, Clayden S, Reyes-Prieto A.** 2015. The Glaucophyta: the blue-green plants in a nutshell. *Acta Societatis Botanicorum Poloniae* **84**, 149–165.
- Jacob A, Lehmann H, Kirst GO, Wiencke C.** 1992. Changes in the ultrastructure of *Prasiola crispa* ssp. *antarctica* under salinity stress. *Botanica Acta* **105**, 41–46.
- Jahns P, Holzwarth AR.** 2012. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta* **1817**, 182–193.
- Jiao C, Sørensen I, Sun X, et al.** 2019. The genome of the charophyte alga *Penium margaritaceum* bears footprints of the evolutionary origins of land plants. *bioRxiv* doi: [10.1101/835561](https://doi.org/10.1101/835561). [Preprint].
- John RP.** 1942. An ecological and taxonomic study of the algae of British soils: I. The distribution of the surface-growing algae. *Annals of Botany* **6**, 323–349.
- Jones VA, Dolan L.** 2012. The evolution of root hairs and rhizoids. *Annals of Botany* **110**, 205–212.
- Ju C, Van de Poel B, Cooper ED, Thierer JH, Gibbons TR, Delwiche CF, Chang C.** 2015. Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nature Plants* **1**, 14004.
- Kaku H, Nishizawa Y, Ishii-Minami N, Akimoto-Tomiya C, Dohmae N, Takio K, Minami E, Shibuya N.** 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences, USA* **103**, 11086–11091.
- Karsten U, Friedl T, Schumann R, Hoyer K, Lembcke S.** 2005. Mycosporine-like amino acids and phylogenies in green algae: *Prasiola* and its relatives from the Trebouxiophyceae (Chlorophyta). *Journal of Phycology* **41**, 557–566.
- Karsten U, Herburger K, Holzinger A.** 2016. Living in biological soil crust communities of African deserts—physiological traits of green algal *Klebsormidium* species (Streptophyta) to cope with desiccation, light and temperature gradients. *Journal of Plant Physiology* **194**, 2–12.
- Karsten U, Lembcke S, Schumann R.** 2007. The effects of ultraviolet radiation on photosynthetic performance, growth and sunscreen compounds in aeroterrestrial biofilm algae isolated from building facades. *Planta* **225**, 991–1000.
- Keeling PJ.** 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annual Review of Plant Biology* **64**, 583–607.
- Kenrick P, Strullu-Derrien C.** 2014. The origin and early evolution of roots. *Plant Physiology* **166**, 570–580.
- Kitzing C, Karsten U.** 2015. Effects of UV radiation on optimum quantum yield and sunscreen contents in members of the genera *Interfilum*, *Klebsormidium*, *Hormidiella* and *Entransia* (Klebsormidiophyceae, Streptophyta). *European Journal of Phycology* **50**, 279–287.
- Knack JJ, Wilcox LW, Delaux PM, Ané JM, Piotrowski MJ, Cook ME, Graham JM, Graham LE.** 2015. Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of microbiota fostered plant colonization of land. *International Journal of Plant Sciences* **176**, 405–420.
- Knopp M, Garg SG, Handrich M, Gould SB.** 2019. Major changes in plastid protein import and the origin of the Chloroplastida. *bioRxiv* doi: [10.1101/799577](https://doi.org/10.1101/799577). [Preprint].
- Kohli A, Sreenivasulu N, Lakshmanan P, Kumar PP.** 2013. The phytohormone crosstalk paradigm takes center stage in understanding how plants respond to abiotic stresses. *Plant Cell Reports* **32**, 945–957.
- Kranner I, Minibayeva FV, Beckett RP, Seal CE.** 2010. What is stress? Concepts, definitions and applications in seed science. *New Phytologist* **188**, 655–673.
- Kranner I, Zorn M, Turk B, Wornik S, Beckett RP, Batič F.** 2003. Biochemical traits of lichens differing in relative desiccation tolerance. *New Phytologist* **160**, 167–176.
- Lang D, Weiche B, Timmerhaus G, Richardt S, Riaño-Pachón DM, Corrêa LG, Reski R, Mueller-Roeber B, Rensing SA.** 2010. Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biology and Evolution* **2**, 488–503.
- Leebens-Mack JH, Barker MS, Carpenter EJ, et al.** 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685.
- Leitz G, Schnepf E, Greulich KO.** 1995. Micromanipulation of statoliths in gravity-sensing *Chara* rhizoids by optical tweezers. *Planta* **197**, 278–288.
- Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O.** 2012. Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences* **31**, 1–46.

- Leverenz RL, Sutter M, Wilson A, et al.** 2015. Photosynthesis. A 12 Å carotenoid translocation in a photoswitch associated with cyanobacterial photoprotection. *Science* **348**, 1463–1466.
- Lewis LA.** 2017. Hold the salt: freshwater origin of primary plastids. *Proceedings of the National Academy of Sciences, USA* **114**, 9759–9760.
- Lewis LA, McCourt RM.** 2004. Green algae and the origin of land plants. *American Journal of Botany* **91**, 1535–1556.
- Li X-P, Björkman O, Shih C, Grossman AR, Rosenquist M, Jansson S, Niyogi KK.** 2000. A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**, 391–395.
- Li X-P, Gilmore AM, Caffarri S, Bassi R, Golan T, Kramer D, Niyogi KK.** 2004. Regulation of photosynthetic light harvesting involves intrathylakoid lumen pH sensing by the PsbS protein. *Journal of Biological Chemistry* **279**, 22866–22874.
- Ligrone R, Duckett JG, Renzaglia KS.** 2000. Conducting tissues and phyletic relationships of bryophytes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**, 795–813.
- Limpens E, Franken C, Smit P, Willemsse J, Bisseling T, Geurts R.** 2003. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302**, 630–633.
- Lind C, Dreyer I, López-Sanjurjo EJ, et al.** 2015. Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure. *Current Biology* **25**, 928–935.
- Madsen EB, Madsen LH, Radutoiu S, et al.** 2003. A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature* **425**, 637–640.
- Mager DM, Thomas AD.** 2011. Extracellular polysaccharides from cyanobacterial soil crusts: a review of their role in dryland soil processes. *Journal of Arid Environments* **75**, 91–97.
- Mähönen AP, Ten Tusscher K, Siligato R, Smetana O, Díaz-Triviño S, Salojärvi J, Wachsman G, Prasad K, Heidstra R, Scheres B.** 2014. PLETHORA gradient formation mechanism separates auxin responses. *Nature* **515**, 125–129.
- Martin WF, Garg S, Zimorski V.** 2015. Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140330.
- Martin-Arevalillo R, Thévenon E, Jégu F, Vinos-Poyo T, Vernoux T, Parcy F, Dumas R.** 2019. Evolution of the auxin response factors from charophyte ancestors. *PLoS Genetics* **15**, e1008400.
- Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, Narusaka Y, Kawakami N, Kaku H, Shibuya N.** 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **104**, 19613–19618.
- Monte I, Ishida S, Zamarreño AM, et al.** 2018. Ligand–receptor co-evolution shaped the jasmonate pathway in land plants. *Nature Chemical Biology* **14**, 480–488.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ.** 2018. The time-scale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **115**, E2274–E2283.
- Müller P, Li XP, Niyogi KK.** 2001. Non-photochemical quenching. A response to excess light energy. *Plant Physiology* **125**, 1558–1566.
- Murik O, Oren N, Shotland Y, Raanan H, Treves H, Kedem I, Keren N, Hagemann M, Pade N, Kaplan A.** 2017. What distinguishes cyanobacteria able to revive after desiccation from those that cannot: the genome aspect. *Environmental Microbiology* **19**, 535–550.
- Mutte SK, Kato H, Rothfels C, Melkonian M, Wong GK-S, Weijers D.** 2018. Origin and evolution of the nuclear auxin response system. *eLife* **7**, 33399.
- Naser V, Shani E.** 2016. Auxin response under osmotic stress. *Plant Molecular Biology* **91**, 661–672.
- Niklas KJ, Cobb ED, Matas AJ.** 2017. The evolution of hydrophobic cell wall biopolymers: from algae to angiosperms. *Journal of Experimental Botany* **68**, 5261–5269.
- Niklas KJ, Dunker AK, Yruela I.** 2018. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *Journal of Experimental Botany* **69**, 1437–1446.
- Nishiyama T, Sakayama H, de Vries J, et al.** 2018. The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464.e24.
- Ohtaka K, Hori K, Kanno Y, Seo M, Ohta H.** 2017. Primitive auxin response without TIR1 and Aux/IAA in the Charophyte alga *Klebsormidium nitens*. *Plant Physiology* **174**, 1621–1632.
- Oldroyd GE.** 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews. Microbiology* **11**, 252–263.
- Oliva J, Rommel S, Fossdal CG, Hietala AM, Nemesio-Goriz M, Solheim H, Elfstrand M.** 2015. Transcriptional responses of Norway spruce (*Picea abies*) inner sapwood against *Heterobasidion parviporum*. *Tree Physiology* **35**, 1007–1015.
- Oliver MJ, Tuba Z, Mishler BD.** 2000. The evolution of vegetative desiccation tolerance in land plants. *Plant Ecology* **151**, 85–100.
- Oren N, Raanan H, Kedem I, Turjeman A, Bronstein M, Kaplan A, Murik O.** 2019. Desert cyanobacteria prepare in advance for dehydration and rewetting: the role of light and temperature sensing. *Molecular Ecology* **28**, 2305–2320.
- Oren N, Raanan H, Murik O, Keren N, Kaplan A.** 2017. Dawn illumination prepares desert cyanobacteria for dehydration. *Current Biology* **27**, R1056–R1057.
- Parfrey LW, Lahr DJ, Knoll AH, Katz LA.** 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences, USA* **108**, 13624–13629.
- Parniske M.** 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Reviews. Microbiology* **6**, 763–775.
- Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK.** 2009. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* **462**, 518–521.
- Pickett-Heaps JD.** 1967. Ultrastructure and differentiation in *Chara* sp. I. Vegetative cells. *Australian Journal of Biological Sciences* **20**, 539–552.
- Pierangelini M, Glaser K, Mikhailyuk T, Karsten U, Holzinger A.** 2019. Light and dehydration but not temperature drive photosynthetic adaptations of basal streptophytes (*Hormidiella*, *Streptosarcina* and *Streptofilum*) living in terrestrial habitats. *Microbial Ecology* **77**, 380–393.
- Pierangelini M, Ryšánek D, Lang I, Adlansnig W, Holzinger A.** 2017. Terrestrial adaptation of green algae *Klebsormidium* and *Zygnema* (Charophyta) involves diversity in photosynthetic traits but not in CO₂ acquisition. *Planta* **246**, 971–986.
- Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D.** 2017. An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology* **27**, 386–391.
- Popper ZA, Michel G, Hervé C, Domozych DS, Willats WG, Tuohy MG, Kloareg B, Stengel DB.** 2011. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annual Review of Plant Biology* **62**, 567–590.
- Pötter E, Kloppstech K.** 1993. Effects of light stress on the expression of early light-inducible proteins in barley. *European Journal of Biochemistry* **214**, 779–786.
- Pressel S, Renzaglia KS, Dicky Clymo RS, Duckett JG.** 2018. Hornwort stomata do not respond actively to exogenous and environmental cues. *Annals of Botany* **122**, 45–57.
- Proteau PJ, Gerwick WH, Garcia-Pichel F, Castenholz R.** 1993. The structure of scytonemin, an ultraviolet sunscreen pigment from the sheaths of cyanobacteria. *Experientia* **49**, 825–829.
- Puttick MN, Morris JL, Williams TA, et al.** 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology* **28**, 733–745.e2.
- Raanan H, Oren N, Treves H, Keren N, Ohad I, Berkowicz SM, Hagemann M, Koch M, Shotland Y, Kaplan A.** 2016. Towards clarifying what distinguishes cyanobacteria able to resurrect after desiccation from those that cannot: the photosynthetic aspect. *Biochimica et Biophysica Acta* **1857**, 715–722.
- Radutoiu S, Madsen LH, Madsen EB, et al.** 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592.
- Raven JA, Edwards D.** 2014. Photosynthesis in early land plants: adapting to the terrestrial environment. In: Hanson DT, Rice SK, eds. *Photosynthesis in bryophytes and early land plants*. Heidelberg: Springer, 29–58.
- Rensing SA.** 2016. (Why) Does evolution favour embryogenesis? *Trends in Plant Science* **21**, 562–573.

- Rensing SA.** 2018. Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology* **42**, 49–54.
- Ridley BL, O'Neill MA, Mohnen D.** 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **57**, 929–967.
- Rippin M, Becker B, Holzinger A.** 2017. Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant & Cell Physiology* **58**, 2067–2084.
- Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH.** 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences, USA* **114**, E7737–E7745.
- Scheibe R, Beck E.** 2011. Drought, desiccation, and oxidative stress. In: Lüttge U, Beck E, Bartels D, eds. *Plant desiccation tolerance*, Vol. 215, *Ecological Studies*. Heidelberg: Springer, 209–232.
- Scheres B, van der Putten WH.** 2017. The plant percepton connects environment to development. *Nature* **543**, 337–345.
- Shinozaki K, Yamaguchi-Shinozaki K, Seki M.** 2003. Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology* **6**, 410–417.
- Shinozawa A, Otake R, Takezawa D, et al.** 2019. SnRK2 protein kinases represent an ancient system in plants for adaptation to a terrestrial environment. *Communications Biology* **2**, 30.
- Skokan R, Medvecká E, Viaene T, et al.** 2019. PIN-driven auxin transport emerged early in streptophyte evolution. *Nature Plants* **5**, 1114–1119.
- Smirnov N.** 1993. The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytologist* **125**, 27–58.
- Smit P, Limpens E, Geurts R, Fedorova E, Dolgikh E, Gough C, Bisseling T.** 2007. *Medicago* LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiology* **145**, 183–191.
- Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F, O'Neill MA, Fei Z, Rose JK, Domozych DS, Willats WG.** 2011. The charophycean green algae provide insights into the early origins of plant cell walls. *The Plant Journal* **68**, 201–211.
- Souffreau C, Vanormelingen P, Sabbe K, Vyverman W.** 2013. Tolerance of resting cells of freshwater and terrestrial benthic diatoms to experimental desiccation and freezing is habitat-dependent. *Phycologia* **52**, 246–255.
- Stancheva R, Hall JD, Herburger K, Lewis LA, McCourt RM, Sheath RG, Holzinger A.** 2014. Phylogenetic position of *Zyggonium ericetorum* (zygnematophyceae, charophyta) from a high alpine habitat and ultrastructural characterization of unusual aplanospores. *Journal of Phycology* **50**, 790–803.
- Stevenson SR, Kamisugi Y, Trinh CH, et al.** 2016. Genetic analysis of *Physcomitrella patens* identifies ABSCISIC ACID NON-RESPONSIVE, a regulator of ABA responses unique to basal land plants and required for desiccation tolerance. *The Plant Cell* **28**, 1310–1327.
- Sun J, Miller JB, Granqvist E, et al.** 2015. Activation of symbiosis signaling by arbuscular mycorrhizal fungi in legumes and rice. *The Plant Cell* **27**, 823–838.
- Sun X, Rikkerink EH, Jones WT, Uversky VN.** 2013. Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *The Plant Cell* **25**, 38–55.
- Sun Y, Harpazi B, Wijerathna-Yapa A, Merilo E, de Vries J, Michaeli D, Gal M, Cuming AC, Kollist H, Mosquna A.** 2019. A ligand-independent origin of abscisic acid perception. *Proceedings of the National Academy of Sciences, USA* **116**, 24892–24899.
- Takahashi S, Murata N.** 2008. How do environmental stresses accelerate photoinhibition? *Trends in Plant Science* **3**, 178–182.
- Tzvetkova-Chevolleau T, Franck F, Alawady AE, Dall'Osto L, Carrière F, Bassi R, Grimm B, Nussaume L, Havaux M.** 2007. The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. *The Plant Journal* **50**, 795–809.
- Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K.** 2010. Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant & Cell Physiology* **51**, 1821–1839.
- Van de Poel B, Cooper ED, Van Der Straeten D, Chang C, Delwiche CF.** 2016. Transcriptome profiling of the green alga *Spirogyra pratensis* (Charophyta) suggests an ancestral role for ethylene in cell wall metabolism, photosynthesis, and abiotic stress responses. *Plant Physiology* **172**, 533–545.
- Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W.** 2012. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *The Plant Cell* **24**, 3506–3529.
- Vogt T.** 2010. Phenylpropanoid biosynthesis. *Molecular Plant* **3**, 2–20.
- Wang S, Li L, Sahu SK, et al.** 2019. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants* doi:10.1038/s41477-019-0560-3
- Weijers D, Wagner D.** 2016. Transcriptional responses to the auxin hormone. *Annual Review of Plant Biology* **67**, 539–574.
- Weng JK.** 2014. The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist* **201**, 1141–1149.
- Wickett NJ, Mirarab S, Nguyen N, et al.** 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* **111**, E4859–E4868.
- Wilhelmsson PKI, Mühlich C, Ullrich KK, Rensing SA.** 2017. Comprehensive genome-wide classification reveals that many plant-specific transcription factors evolved in streptophyte algae. *Genome Biology and Evolution* **9**, 3384–3397.
- Wolf L, Rizzini L, Stracke R, Ulm R, Rensing SA.** 2010. The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-B radiation. *Plant Physiology* **153**, 1123–1134.
- Xu B, Ohtani M, Yamaguchi M, et al.** 2014. Contribution of NAC transcription factors to plant adaptation to land. *Science* **343**, 1505–1508.
- Xu D, Duan X, Wang B, Hong B, Ho T, Wu R.** 1996. Expression of a late embryogenesis abundant protein gene, *HVA1*, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiology* **110**, 249–257.
- Yoshida K, Inoue N, Sonobe S, Shimmen T.** 2003. Involvement of microtubules in rhizoid differentiation of *Spirogyra* species. *Protoplasma* **221**, 227–235.
- Yoshida K, Shimmen T.** 2009. Involvement of actin filaments in rhizoid morphogenesis of *Spirogyra*. *Physiologia Plantarum* **135**, 98–107.
- Yoshida T, Mogami J, Yamaguchi-Shinozaki K.** 2014. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Current Opinion in Plant Biology* **21**, 133–139.
- Żabka A, Polit JT, Winnicki K, Paciorek P, Juszcak J, Nowak M, Maszewski J.** 2016. PIN2-like proteins may contribute to the regulation of morphogenetic processes during spermatogenesis in *Chara vulgaris*. *Plant Cell Reports* **35**, 1655–1669.
- Zeng Q, Chen X, Wood AJ.** 2002. Two early light-inducible protein (ELIP) cDNAs from the resurrection plant *Tortula ruralis* are differentially expressed in response to desiccation, rehydration, salinity, and high light. *Journal of Experimental Botany* **53**, 1197–1205.
- Zhang Y, Launay H, Schramm A, Lebrun R, Gontero B.** 2018. Exploring intrinsically disordered proteins in *Chlamydomonas reinhardtii*. *Scientific Reports* **8**, 1–11.
- Zhang B, Liu J, Yang ZE, Chen EY, Zhang CJ, Zhang XY, Li FG.** 2018. Genome-wide analysis of GRAS transcription factor gene family in *Gossypium hirsutum* L. *BMC Genomics* **19**, 348.
- Zhao C, Wang Y, Chan KX, et al.** 2019. Evolution of chloroplast retrograde signaling facilitates green plant adaptation to land. *Proceedings of the National Academy of Sciences, USA* **116**, 5015–5020.
- Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA.** 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Molecular Genetics and Genomics* **278**, 393–402.
- Zorn M, Pfeifhofer HW, Grill D, Kranner I.** 2001. Responses of plastid pigments to desiccation and rehydration in the desert lichen *Ramalina maciformis*. *Symbiosis* **31**, 201–211.

3.2 Publication II: Submergence of the filamentous Zygnematophyceae *Mougeotia* induces differential gene expression patterns associated with core metabolism and photosynthesis

This research paper was published online in the Journal “Protoplasma” in December 2021. The full article as well as supplementary figures and supplementary datasets can be found online:

<https://doi.org/10.1007/s00709-021-01730-1>

Contribution of Janine Fürst-Jansen, first author

J. M. R. Fürst-Jansen performed all the culturing for the physiological experiments. She generated all the microscope pictures for the alga strain that stands in the focus of this manuscript. She designed together with J.d.V. and planned the (photo-)physiological setup, the transition times and cycles, and performed all chlorophyll a fluorescence measurements. Together with J.d.V she performed computational analysis of the RNA sequencing data. To process this large transcriptomic data set, J. M. R. Fürst-Jansen performed statistical analysis. After filtering out intriguing gene candidates from this data set, she performed a phylogenetic analysis together with J.d.V. She designed and organized figure 3 and figure 4 and helped to process figure 5. She wrote parts of the manuscript and critically read and revised the manuscript.



Submergence of the filamentous Zygnematophyceae *Mougeotia* induces differential gene expression patterns associated with core metabolism and photosynthesis

Janine M.R. Fürst-Jansen¹ · Sophie de Vries¹ · Maike Lorenz² · Klaus von Schwartzberg³ · John M. Archibald⁴ · Jan de Vries^{1,5,6}

Received: 16 June 2021 / Accepted: 6 December 2021 / Published online: 22 December 2021

© The Author(s) 2021

Abstract

The streptophyte algal class Zygnematophyceae is the closest algal sister lineage to land plants. In nature, Zygnematophyceae can grow in both terrestrial and freshwater habitats and how they do this is an important unanswered question. Here, we studied what happens to the zygnematophyceae alga *Mougeotia* sp., which usually occurs in permanent and temporary freshwater bodies, when it is shifted to liquid growth conditions after growth on a solid substrate. Using global differential gene expression profiling, we identified changes in the core metabolism of the organism interlinked with photosynthesis; the latter went hand in hand with measurable impact on the photophysiology as assessed via pulse amplitude modulation (PAM) fluorometry. Our data reveal a pronounced change in the overall physiology of the alga after submergence and pinpoint candidate genes that play a role. These results provide insight into the importance of photophysiological readjustment when filamentous Zygnematophyceae transition between terrestrial and aquatic habitats.

Keywords Streptophyte algae · Charophytes · RNAseq · Algal culturing · Algal physiology

Handling Editor: Andreas Holzinger

✉ Jan de Vries
devries.jan@uni-goettingen.de

Janine M.R. Fürst-Jansen
Janine.fuerst-jansen@uni-goettingen.de

Sophie de Vries
sophie.devries@uni-goettingen.de

Maike Lorenz
mlorenz@uni-goettingen.de

Klaus von Schwartzberg
Klaus.von.Schwartzberg@uni-hamburg.de

John M. Archibald
john.archibald@dal.ca

Science, University of Goettingen, Nikolausberger Weg 18,
37073 Goettingen, Germany

³ Institute of Plant Science and Microbiology, Microalgae and Zygnematophyceae Collection Hamburg (MZCH) and Aquatic Ecophysiology and Phycology, Universität Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany

⁴ Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, NS B3H 4R2, Canada

⁵ Goettingen Center for Molecular Biosciences (GZMB), University of Goettingen, Justus-von-Liebig-Weg 11, 37077 Goettingen, Germany

⁶ Campus Institute Data Science (CIDAS), University of Goettingen, Goldschmidtstr. 1, 37077 Goettingen, Germany

¹ Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Goettingen, Goldschmidtstr. 1, University of Goettingen, 37077 Goettingen, Germany

² Department of Experimental Phycology and SAG Culture Collection of Algae, Albrecht-von-Haller Institute for Plant

Introduction

Streptophyte algae diverged from the chlorophytes and prasinodermophytes between 700 and 1000 million years ago (Zimmer et al. 2007; Morris et al. 2018; Li et al. 2020). They form a paraphylum that is sister to the monophyletic Embryophyta, the land plants—together, land plants and streptophyte algae form the monophylum Streptophyta (Wickett et al. 2014). One of the most important questions in the field of land plant evolution is which particular lineage of streptophyte algae within this paraphylum represents the sister lineage to land plants. Streptophyte algae encompass a diverse range of organisms, including the unicellular Mesostigmatophyceae and Chlorokybophyceae (cell packages), consisting of only a few species (see also Irisarri et al. 2021), the unicellular and filamentous Klebsormidiophyceae (Mikhailyuk et al. 2015), and the streptophyte algae within Phragmoplastophyta that include morphologically complex multicellular organisms such as the Charophyceae—and the land plants. Various lines of evidence indicate that, among these Phragmoplastophyta, the Zygnematophyceae represent the sister lineage to land plants (Wodniok et al. 2011; Wickett et al. 2014; Leebens-Mack et al. 2019). It is thus of considerable interest what physiological properties these organisms possess—combined with data on land plants, such an understanding makes it possible to infer the physiology of the earliest land plants (Fürst-Jansen et al. 2020).

A key piece of the puzzle of understanding plant terrestrialization is the difference between growth in an aquatic environment and growth in a terrestrial habitat with limited water supply. Throughout the course of evolution, various algal lineages have mastered this so-called wet-to-dry transition. This is no small feat. The terrestrial habitat poses various challenges for a photosynthesizing organism, including fluctuations in abiotic factors such as temperature, water availability, and intensity and quality of irradiance (Foyer et al. 1994; Karsten et al. 2007; Holzinger et al. 2014; Ohama et al. 2017).

Terrestrial algae meet the challenges of their habitat with various physiological adaptations (Holzinger and Pichrtová 2016). These include the presence of mycosporine-like amino acids (MAAs) found in both chlorophyte and streptophyte algae. MAAs have UV-protecting properties. Among streptophyte algae, the Klebsormidiophyceae *Hormidiella* and *Klebsormidium* stand out by producing potent sunscreen MAAs with an absorption maximum at 325 and 324 nm (Kitzing and Karsten 2015). While such MAAs have not been reported for Zygnematophyceae, *Zygnema* spp. are known to produce phenolic compounds upon elevated UV irradiance (Pichrtová et al. 2013). Indeed, the unicellular Zygnematophyceae *Penium*

margaritaceum was recently reported to contain flavonoids (Jiao et al. 2020). While the exact biochemical routes towards these metabolites are currently elusive, homologs of genes coding for core enzymatic biosynthetic steps that lead to relevant precursor metabolites in land plants (the phenylpropanoid pathway *sensu lato*) are also found in streptophyte algae (de Vries et al. 2017, 2021). Recently, Renault et al. (2019) highlighted the putative links between phenylpropanoid biosynthesis in streptophyte algae and shared ancestral chassis for producing hydrophobic polymers from which lignin, cutin, suberin, and sporopollenin arose. Indeed, Zygnematophyceae surround their zygotes with resistant polymers resembling sporopollenin (de Vries et al. 1983; Poulícková et al. 2007). Recently, Permann et al. (2021) employed glycan labeling as well as Raman spectroscopy to zygospores of *Mougeotia disjuncta* (which belongs to the same algal genus as the strains analyzed here); they found these zygospores to consist of a combination of carbohydrates, lipids, and aromatic compounds, speaking to sporopollenin-like material.

UV irradiance is not the only sunlight-associated challenge in the terrestrial habitat. Photosynthetically active radiance (PAR) reaches much higher levels on the surface of the earth as opposed to an aquatic environment, where the sunlight is buffered by the absorptive properties of water. One of the main mechanisms that mitigates damage to the components of the photosynthetic light reaction, in particular the vulnerable photosystem II, is non-photochemical quenching (NPQ; Müller et al. 2001; Jahns and Holzwarth 2012). The first and fastest response of NPQ is energy-dependent quenching (qE). Its activation hinges upon conformational changes in the photosystem and the detection of an altered pH in the thylakoid lumen (Krause et al. 1982). While their predominance varies across the green lineage, evidence suggests that the LHCSR (light-harvesting complex stress-related protein) and/or PSBS (photosystem II subunit S) proteins play a major role in this process (Li et al. 2000; Peers et al. 2009; Gerotto and Morosinotto 2013; Correa-Galvis et al. 2016). It is nevertheless prudent to note that some chlorophyte algae seem to lack qE (Christa et al. 2017). The result of NPQ is that superfluous energy, which cannot be meaningfully channeled into the light reaction chain, dissipates as harmless heat.

The role of NPQ and acclimation processes of the photosystem has been extensively studied in terrestrial streptophyte algae. For example, Herburger and Holzinger (2015) found that the photosynthetic effective quantum yield is strongly reduced in *Klebsormidium* strains upon desiccation but also recovers fully upon rehydration suggesting a high desiccation tolerance. Furthermore, Karsten et al. (2014) found that the sister group of *Klebsormidium*, *Interfilum*, also appears to have similar characteristics regarding high tolerance to stressors such as desiccation but also

temperature that reflect in their photosynthetic physiology. That said, not only the family of Klebsormidiaceae shows this high tolerance to stressors. In the class of Zygnematophyceae, Holzinger et al. (2018) found that after UV-treatment in different *Zygnema* strains their effective quantum yield recovers completely in some cases. There are however other conserved mechanisms for photoprotection acting in algae. One is the expression of EARLY LIGHT INDUCED PROTEIN (ELIP). ELIPs are chlorophyll *a/b*-binding proteins that accumulate under stress and have a photoprotective function (Montané et al. 1997; Hutin et al. 2003). Elevated expression of ELIP-coding genes under light and temperature stress has now been reported for the Zygnematophyceae *Zygnema* and *Mougeotia* (de Vries et al. 2018; Rippin et al. 2019; de Vries et al. 2020). As with the relevance of NPQ under water scarcity, *ELIP* expression is also induced in desiccated *Zygnema* (Rippin et al. 2017). Thus, while we know about physiological responses of Zygnematophyceae challenged with water scarcity, we know very little about the reverse process—which is of similar importance for organisms that thrive in temporary water bodies. Plant terrestrialization likely entailed a repetition of several wet-to-dry and dry-to-wet transitions; therefore, investigating both transitions is important. Furthermore, living on land means a steady change between wet and dry conditions (rain, fog, and dew). *Mougeotia* spp. live in a variety of freshwater habitats, many of them are temporary habitats such as ditches and small temporary ponds.

In this study, we have used a laboratory controlled environmental shift approach to emulate what happens to the filamentous zygnematophycean alga *Mougeotia* sp., which predominantly lives in freshwater habitats, shortly after being submerged. Our data highlight photosynthesis-associated physiological responses and the global gene expression patterns that bring them to bear.

Material and methods

Culturing and treatment

For the RNAseq experiments, *Mougeotia* sp. MZCH 240 (which we obtained from the Microalgae and Zygnematophyceae Collection, Hamburg, Germany, [von Schwartzberg et al. 2013]) was cultured as described in de Vries et al. (2020)—algae were grown for 7 days on modified freshwater F/2 (Guillard 1975) with 1% agar at 22°C and 120 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ from an LED light source (12h/12h light/dark cycle) in 9 cm plates. For submergence, 10 mL of temperature-adjusted liquid freshwater F/2 (Guillard 1975) were added to each agar plate; for RNAseq, algae were directly transferred into Trizol (Thermo Fisher, Waltham, MA, USA) after 4 h of submergence.

For the photophysiological experiments, *Mougeotia scalaris* strain SAG 164.80 (of the Culture Collection of Algae, Göttingen, Germany; Friedl and Lorenz 2012) and *Mougeotia* sp. MZCH 240 were grown on (i) WHM medium (*M. scalaris* SAG 164.80; Nichols 1973) with 1% agar or in liquid WHM medium or (ii) modified freshwater F/2 (*Mougeotia* sp. MZCH 240) with 1% agar or in liquid F/2 medium at 22°C and 120 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ from an LED light source (12h/12h light/dark cycle). For submergence, 10 mL of temperature-adjusted liquid WHM (*M. scalaris* SAG 164.80) or liquid F/2 (*Mougeotia* sp. MZCH 240) were added to each plate and F_v/F_m was measured after various incubation timepoints (2h, 4h, 6h, 8h, 24h; plus 1h and 3h for *Mougeotia scalaris* SAG 164.80). For morphological observations and micrographs, *M. scalaris* SAG 164.80 and *Mougeotia* sp. SAG 650-1 were used as additional comparative material and grown either in liquid or solid for 9 weeks on Desmidiacean Medium (MiEB12; medium 7 of Schlösser 1994). For *Mougeotia* sp. MZCH 240, microscope pictures were taken after the 24h timepoint under the growing conditions described above.

RNA extraction and sequencing

RNA extraction and sequencing procedures were described in de Vries et al. (2020). In brief, we extracted RNA in six biological replicates from the control samples and in biological triplicates from the liquid-treated samples. For RNA extraction, algae were directly transferred into 1 mL of Trizol using a sterilized spatula (Thermo Fisher, Waltham, MA, USA); extraction procedures were carried out in accordance to the protocol provided by the manufacturer. Isolated RNA was treated with DNase I (Thermo Fisher), quality assessed on a formamide agarose gel, quantified using a Nanodrop spectrometer (Thermo Fisher), and shipped to Genome Québec (Montreal, Canada) for sequencing. There, RNA was quality checked again, using a Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA). Libraries were constructed using the NEB mRNA stranded Library preparation kit (New England Biolabs, Beverly, MA, USA), on the Illumina NovaSeq6000 platform.

RNAseq analyses: data processing, statistics, KEGG, and GOterm

Initial processing of the RNAseq data was described in de Vries et al. (2020). In brief, reads were checked for quality using FASTQC version 0.11.7 (FASTQC 2018), trimmed with TRIMMOMATIC v0.36 (Bolger et al. 2014; settings: ILLUMINACLIP:TruSeq3-PE- 2.fa:2:30:10:2:TRUE HEADCROP:10 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36), and quality checked again using FASTQC v0.11.7. For details on read data, see the “Data availability”

section. The transcriptome assembly using the TRINITY pipeline (Haas et al. 2013), RSEM (RNA-Seq by Expectation Maximization)-based read mapping (Li and Dewey 2011) was carried out and described in de Vries et al. (2020).

Negative binomial distribution-based statistical analyses of the read counts were performed using edgeR version 3.28.0 (Robinson et al. 2010), taking the biological triplicates into account. For all downstream analyses, only gene expression changes with a Benjamini-Hochberg-corrected p value ≤ 0.001 and significantly elevated differential gene expression (\log_2 (fold change) ≤ -1 or \log_2 (fold change) ≥ 1) were considered.

For gene expression analysis based on KEGG orthologs, we worked with expression levels in TPM that were normalized via TMM (trimmed mean of M values; Robinson and Oshlack 2010). These data against KEGG pathways occur in land plants. If multiple transcripts had the same KEGG ortholog as their best hit, their expression values were combined—for the final calculations, a given KEGG ortholog had one TMM-normalized TPM value.

For GO term enrichment using GOrilla (Eden et al. 2009), we used AGI numbers obtained by querying the predicted *Mougeotia* proteins against Arabidopsis in a BLASTp in a comparison of two unranked list of genes. For this, we used all obtained Arabidopsis homologs (i.e. the best BLASTp hits) as the background set (as the whole transcriptome) and all significantly regulated genes as target set—one target set for all up-regulated genes, one for all downregulated genes.

Photophysiology

All measurements of the maximum-quantum yield (F_v/F_m) were done using the maxi version of the Imaging-PAM (ImagMAX/L, M-series, Walz) with an IMAG-K5 CCD camera controlled with the ImagingWinGigE (V2.32) software. Treated as well as control samples were dark adapted 10–30 min before measurement. For F_v/F_m measurement, a short saturation pulse with intensity 10 (setup 1-3; level 3 for setup 4) was applied, which is the standard intensity for the IMAGING-PAM. Within the four experimental setups (three with SAG 164.80 and a fourth with MZCH 240), the settings for measuring light and gain were adjusted slightly (setup 1: measuring light 4, gain 2; setup 2: measuring light 1, gain 10; setup 3(+setup 4): measuring light 1, gain 3). A special SP-Routine was not applied to modify the signal to noise ratio of the fluorescence measurement. Statistical analysis was done using Mann-Whitney U tests (Mann and Whitney 1947) with R (version 3.6.1).

Phylogenetic analysis

To explore whether the ABA3 and PAP homolog we detected in the RNA-Seq-based de novo assembly represents

an ABA3 ortholog in *Mougeotia* sp. MZCH 240, we used BLASTp to mine the protein datasets of (i) the land plants *Anthoceros agrestis* (Li et al. 2020), *Arabidopsis thaliana* (Lamesch et al. 2012), *Azolla filiculoides* (Li et al. 2018), *Marchantia polymorpha* (Bowman et al. 2017), *Nicotiana tabacum* (Sierro et al. 2014), *Physcomitrium patens* (Lang et al. 2018), *Selaginella moellendorffii* (Banks et al. 2011); (ii) the streptophyte algae *Chlorokybus atmophyticus* CCAC 0220 (Wang et al. 2020), *Chara braunii* S 276 and S 277 (Nishiyama et al. 2018), *Klebsormidium nitens* NIES-2285 (Hori et al. 2014), *Mesotaenium endlicherianum* SAG 12.97 (Cheng et al. 2019), *Mesostigma viride* CCAC 1140 (Wang et al. 2020), *Spirogloea muscicola* CCAC 0214 (Cheng et al. 2019); (iii) *Bathycoccus prasinus* RCC 1105 (Moreau et al. 2012), *Chlamydomonas reinhardtii* CC-503 (Merchant et al. 2007), *Volvox carteri* f. *nagariensis*, Eve (Prochnik et al. 2010).

All obtained sequences were aligned using MAFFT (Katoh and Standley 2013) with the L-INS-I settings. The alignment was used for computing a maximum likelihood phylogeny using IQ-TREE multicore v.1.5.5 for Linux 64-bit built (Nguyen et al. 2015) with 100 bootstrap replicates; the best model for protein evolution (WAG+F+I+G4 for ABA3 and WAG+I+G4 for PAP; both were chosen according to Bayesian Information Criterion) was determined using ModelFinder (Kalyaanamoorthy et al. 2017).

Results and discussion

Submergence in liquid medium triggers the differential expression of core metabolism and photosynthesis-related genes in *Mougeotia* sp.

Using the filamentous zygneematophycean alga *Mougeotia* sp. (a representative species of the zygneematophycean clade), we analyzed differences in the transcriptome of *Mougeotia* sp. MZCH 240 under two growth conditions: (i) growth on solid medium and (ii) 4 h after submergence with liquid medium.

Using the Illumina NovaSeq 6000 platform (operated by Genome Quebec), we obtained ~159 million paired reads for the solid growth condition (6 biological replicates) and 100 million paired reads for the sample taken 4h after submergence (3 biological replicates). After quality checking and trimming, we mapped these reads onto the transcriptome assembly of *Mougeotia* sp. MZCH 240 (de Vries et al. 2020) using the RSEM toolkit included in the TRINITY pipeline. Using this transcriptome assembly, we worked with 4961 genes, of which 438 genes showed more than 2-fold upregulation and 775 genes showed more than 2-fold downregulation (Figure 1A; more on statistic scrutinization below).

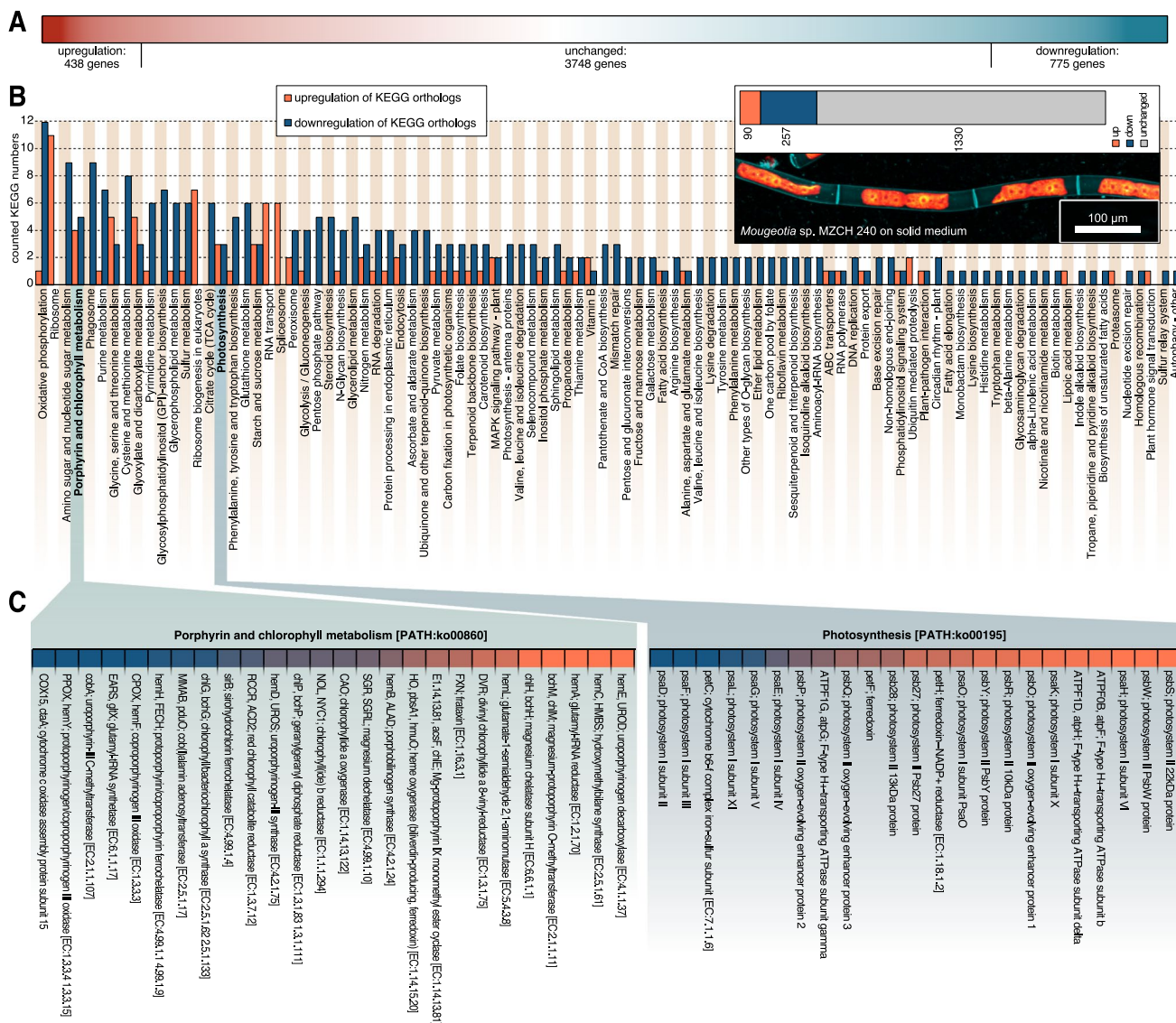
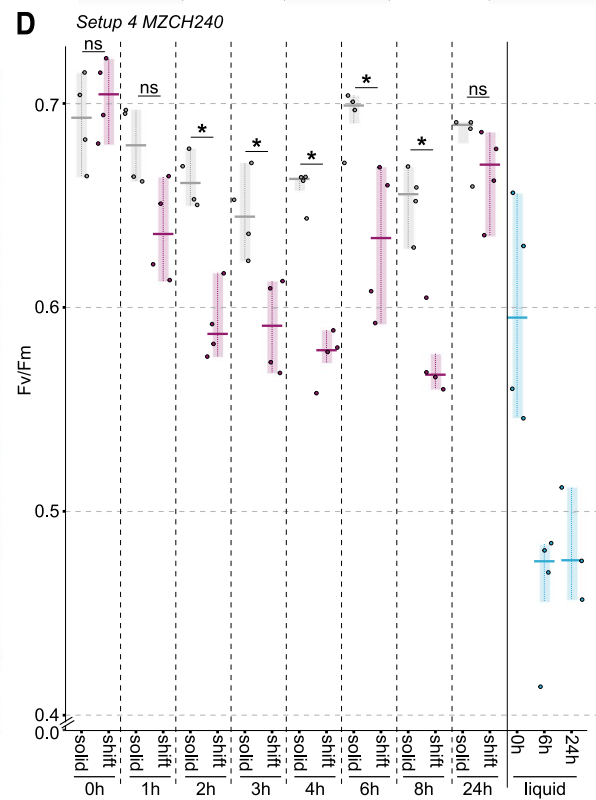
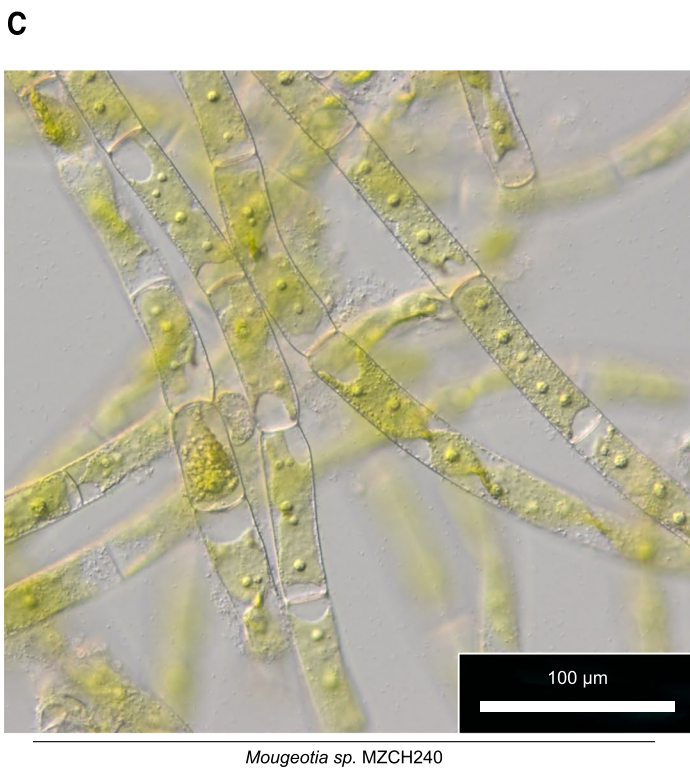
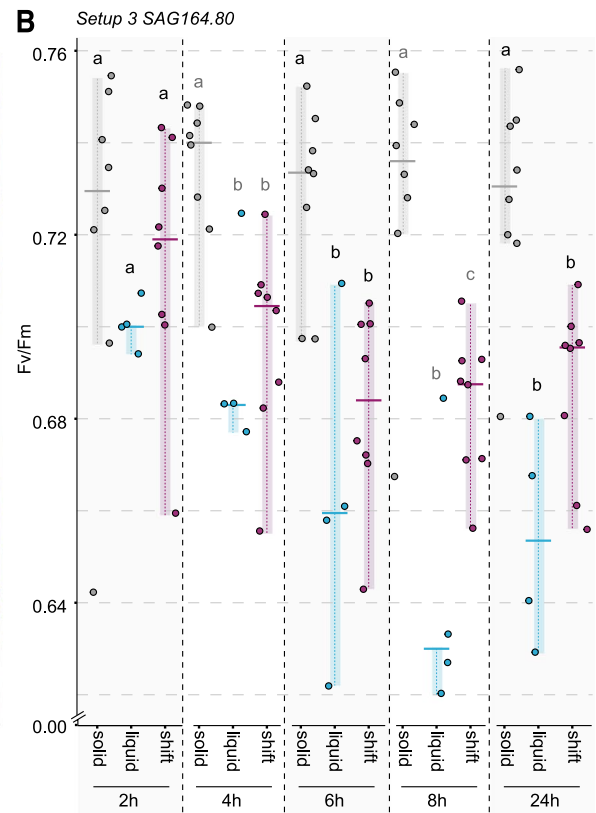
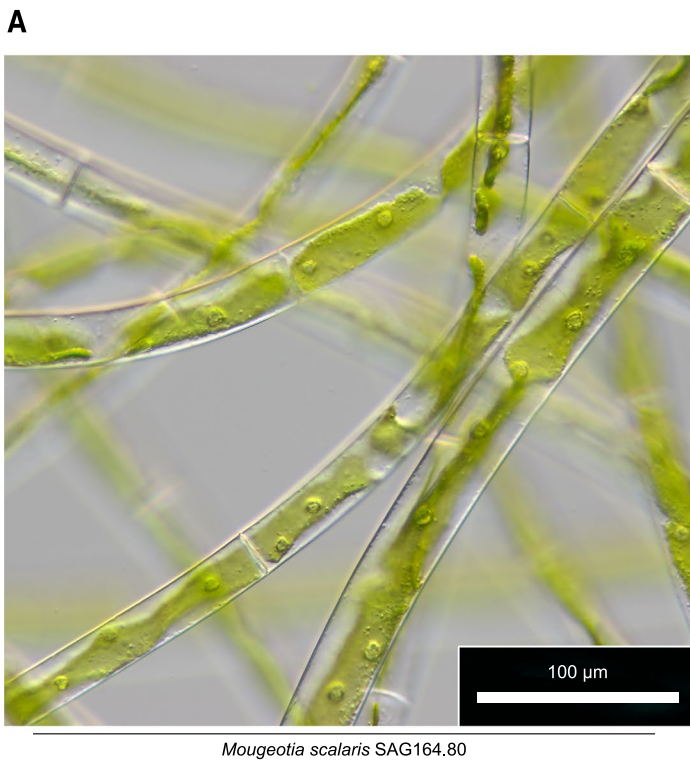


Fig. 1 Global gene expression patterns in *Mougeotia* sp. MZCH 240. **A** Gradient-colored depiction (red up-regulated, white unchanged, and blue downregulated genes) of the differential global gene expression profile of all 4961 genes analysed in this study; the differential responses were obtained by comparing global gene expression of *Mougeotia* sp. MZCH 240 cultured on solid medium and submerged for 4h versus control (growth on solid medium). **B** Gene expression pattern of various KEGG orthologs in *Mougeotia* sp. MZCH 240. Biological replicates (at least triplicates) of gene expression data (TPM_{TMM-normalized}) were summed up and set relative to the control condition data (submergence/control) and then mapped against the Kyoto Encyclopedia of Genes and Genomes (KEGG). An up- or downregulation of a KEGG ortholog was considered if it had a ≥ 2 -fold change in gene expression levels. A bar diagram depicts the numbers of all up- (orange) or downregulated (dark blue) KEGG orthologs in the 118 detected KEGG plant pathways in *Mougeotia* sp. MZCH 240 4h after being submerged (shift) in liquid medium

First, we were interested in getting an overview over transcriptomic differences induced by submergence in liquid

compared to the control culture, which was kept on solid medium. On the upper right side all counted KEGG numbers from up- (90) or down- (257) regulated KEGG orthologs are shown in a stacked bar plot together with 1330 KEGG orthologs with unchanged (grey) gene expression patterns; below is a confocal micrograph of *Mougeotia* sp. MZCH 240 under control conditions (grown in modified freshwater F/2 with 1% agar 22°C and 120 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$)—cell walls were made visible using 1% calcofluor white staining (teal false colored), the plastids are shown in a false-colored red-orange gradient based on their chlorophyll *a* autofluorescence. **C** A heatmap of the gene expression patterns in *Mougeotia* sp. MZCH 240 of the two KEGG plant pathways “Porphyrin and chlorophyll metabolism [PATH:ko00860]” and “Photosynthesis [PATH:ko00195]” in detail. Data is shown as \log_2 (fold change_{submergence/control}) in a color gradient ranging from dark blue (downregulation) to orange (upregulation). Unchanged expression levels are not depicted here

medium; we used the pathway framework of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.



We used BLASTKOALA (Kanehisa et al. 2016) to identify KEGG orthologs among our de novo assembled transcripts and then linked the expression values (fold change) to the

corresponding KEGG numbers. All gene expression values for a given KEGG ortholog were summed up as described in de Vries et al. (2020). A KEGG ortholog was considered

Fig. 2 Plastid morphology and photophysiological characteristics (F_v/F_m) in *Mougeotia scalaris* SAG 164.80 and *Mougeotia* sp. MZCH 240. **A** Light micrograph of *M. scalaris* SAG 164.80 in liquid medium. **B** Maximum PSII quantum yield (F_v/F_m) in *M. scalaris* SAG 164.80 solid- and liquid-medium control samples (grown for 7 days on WHM-Medium at 20°C, 120 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$) as well as samples treated with the liquid shift—which were grown on solid medium and submerged in 10 ml liquid medium. **C** Light micrograph of *Mougeotia* sp. MZCH 240 24h after submergence. **D** F_v/F_m values for *Mougeotia* sp. MZCH 240 when grown on F/2 medium for 7 days at 22°C, 120 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ on solid and liquid medium. Liquid shift was achieved by adding 10 ml liquid medium to algal cultures grown on solid medium. F_v/F_m values were collected at 0, 1, 2, 3, 4, 6, 8, and 24 h after the shift and for the control on solid medium. Owing to the low growth rate in liquid medium values for F_v/F_m were measured only at 0, 6, and 24 h for liquid cultures of *Mougeotia* sp. MZCH 240. F_v/F_m for **B** and **D** was measured from the same sample at several time points (from 2h up to 24h) after liquid medium was added by using an IMAGMAX/L PAM with an IMAG-K5 CCD camera (for details, see the “Material and methods” section). Solid control samples are depicted in grey, liquid control samples are shown in blue, and liquid-treated samples (shift) are depicted in pink. Statistical analysis was done using Mann-Whitney *U* tests with R (version 3.6.1); significant differences at $p < 0.05$ are depicted using letters and asterisks

up- or downregulated if it had a ≥ 2 -fold change in gene expression level. 118 KEGG pathways were identified (Figure 1B). In total, expression values for 1677 KEGG orthologs (corresponding 1176 unique KEGG orthologs) were mapped across pathways, among which 90 orthologs were up-regulated and 257 downregulated in *Mougeotia* sp. MZCH 240 after the shift to liquid conditions; this adds up to a total of 347 responsive KEGG orthologs while 1330 orthologs showed an unchanged response (see the overview in the top right section of Figure 1B).

Most prominent among the top 20 most responsive KEGG pathways were those associated with core metabolic processes such as “oxidative phosphorylation [PATH:ko00190]”, “ribosome [PATH:ko03010]”, and “amino sugar and nucleotide sugar metabolism [PATH:ko00520]” with 13, 11, and 9 differentially regulated KEGG orthologs respectively. We interpret categories such as ribosome, nucleotide metabolism, and any amino acid metabolism as a readout often observed upon any treatment/shift in environmental conditions: the basal molecular machineries of the cells are responding: they power up for making a range of new/different proteins, resulting in a need to produce a different set of amino acids for making these; prior, as well as alongside of this, they make, process, and transport RNA. Similarly, the downregulation of respiration (oxidative phosphorylation and the citrate cycle) can likely be traced to an overall impacted metabolism. We hence searched whether the data speak to any such process upstream and honed in on photosynthesis—the source of carbon for any photoautotroph.

Two photosynthesis-related pathways, namely, “Porphyrin and chlorophyll metabolism [PATH:ko00860]” (4th most responsive, when considering both up- and downregulated KEGG orthologs) and “Photosynthesis [PATH:ko00195]” (16th most responsive), contained some of the most highly differentially regulated KEGG orthologs among all 118 pathways; with 4 up- and 5 downregulated KEGG orthologs for the Porphyrin and chlorophyll metabolism pathway and 3 up- and 3 downregulated KEGG orthologs for the photosynthesis pathway (Figure 1C). The finding of photosynthesis-associated genes might explain why other pathways of core metabolism, as well as housekeeping genes are also affected—photosynthesis is at the heart of plant and algal physiology. If the primary fixation of carbon mediated by photosynthesis is affected by a changing environment, it is conceivable that other pathways dependent on the fixed carbon tag along.

Submergence in liquid medium impacts the photophysiology of two strains of *Mougeotia*

The top three up- and downregulated KEGG orthologs that belong to the pathway “photosynthesis” mainly fall into the category of photosystem I and II subunits, which suggests pronounced readjustment of the composition and stoichiometry of main components that form the chain of proteins acting in the photosynthetic light reaction; this likely goes hand in hand with selectively elevated turnover rates. We thus honed in on the plastid-associated biology of *Mougeotia*. For this, we made use of the emerging model system *M. scalaris* SAG 164.80 (Regensdorff et al. 2018; Figure 2A) and investigated its photophysiological changes after submergence using PAM. For this, we used three experimental setups, each with a minimum of three replicates. In a first experimental setup, we tested changes in maximum quantum yield (F_v/F_m) over time when *M. scalaris* was grown on plates and in liquid culture. We initially explored whether photophysiological changes occur over a short period (4h, Figure S1A; setup 1) during daily growth; in the second setup, we investigated whether there are differences in daily performance (24h, Figure S1B; setup 2). On solid medium, F_v/F_m appeared stable when measurements were only 4h apart, yet when tested daily, we found a decrease in F_v/F_m in the algal culture (p value = 0.029; Figure S1A and B). In liquid culture, F_v/F_m increased from 0.382 ± 0.020 to 0.412 ± 0.018 after 4h (p value = 0.041) in setup 1 but was similar to the starting value after 24h in setup 2 (0.613 ± 0.017 to 0.632 ± 0.015 , p value = 0.0343; Figure S1A and B). We, however, noted that F_v/F_m of *M. scalaris* SAG 164.80 differed significantly at the first measurement (solid 0h: p value = 0.029; liquid 0h: p value = 0.0095).

Despite differences in the actual values of F_v/F_m in the algal culture, we observed a similar trend after submergence of the algae on plate. Short after submergence (1h), F_v/F_m was similar to that of algal culture grown on non-submerged plates. That said, over time, we saw a decrease of F_v/F_m that significantly differs from that of algae grown on agar after 4h (Figure S1A and B). It is noteworthy, however, that the values between liquid culture, solid culture and the submerged culture are similar at 24h (Figure S1B). The data thus remained inconclusive because only two time points were sampled for liquid- and solid-grown algae and the time points were taken from different cultures.

In a next step, we (i) traced the photophysiological properties of the same liquid-grown, solid-grown and submerged algal cultures over time and (ii) compared the differences in F_v/F_m between the different growth conditions (setup 3; Figure 2B) at a given time point. Both solid and liquid grown cultures remained steady over time in their F_v/F_m (Table 1). In contrast, the submerged cultures tend to have a significantly decreased F_v/F_m after 6, 8, and 24h compared to the F_v/F_m at 2h. This agrees with the decreasing trend observed for F_v/F_m in the first two experiments, where different cultures were measured at the different time points. Additionally, this shows that while the decrease in F_v/F_m for the submergence was real, the differences between the different time points for cultures grown in liquid or on solid medium stems from fluctuations in cultures and culturing.

We next compared the data from a given time point between the different growth conditions. While F_v/F_m did not differ at 2h, it was always higher in solid grown medium than in liquid and submerged cultures from 4h onwards (Figure 2B). Liquid and submerged cultures showed mainly

similar F_v/F_m values, the only exception being 8h after treatment; at this time point, the liquid cultures had a significantly lower F_v/F_m than the submerged culture. Taken together, our data suggest that submerged cultures behave—after an initial equilibration phase—more similar to cultures grown in liquid medium than on solid medium. While the trend is largely reproducible, cultural fluctuations in initial photosystem performance nevertheless exist.

In order to scrutinize whether the observations we made on *Mougeotia scalaris* SAG 164.80 (Figure 2A and B) also hold for the strain on which the transcriptomic analyses were performed, we carried out the PAM-based investigations with *Mougeotia* sp. MZCH 240. The cultures of MZCH 240 had F_v/F_m values at the start of the experiment that were (a) similar for the cultures (grown on solid 1% agar medium) that were about to be submerged (shift) and those that were kept as the untreated control (solid) (0.703 ± 0.017 (shift) and 0.691 ± 0.020 (solid), no significant difference) and (b) comparable to the values of the strain SAG164.80. Cultures of MZCH 240 grown in liquid medium generally had lower F_v/F_m values [0.598 ± 0.053 (t_0), 0.462 ± 0.033 (6h), and 0.482 ± 0.028 (24h)]. Already after 2h, submerged cultures had significantly ($p=0.029$) lower F_v/F_m values; this trend of significantly lower ($p < 0.05$) F_v/F_m values continued at time points 3h, 4h, 6h, and 8h. After 24h, the submerged cultures appeared to have acclimated to their new culturing conditions as the F_v/F_m values were almost back to t_0 : 0.665 ± 0.019 (shift) and 0.682 ± 0.013 (solid)—with no significant difference. This is in contrast to the physiological behavior of SAG164.80, which did not acclimate to submergence within a 24h timeframe. Regardless, it should be re-iterated that MZCH 240 showed significantly lower F_v/F_m values at 4h after submergence, which is the time point that was used for transcriptome analyses of this strain; both MZCH 240 and SAG164.80 behaved alike at this time point with regard to their photophysiology assessed through F_v/F_m .

While the photophysiology had recovered at 24h after submergence, only then did morphological differences between the solid control and submerged cultures emerge in *Mougeotia* sp. MZCH 240. The shifted cultures more readily accumulated storage granules (Figure 3); whether these might speak to lipid droplets, as potentially occurring in *Spirogyra* (see also de Vries and Ischebeck 2020), is unclear. Such granules were sometimes also found in samples of the solid control group. However, the most notable phenotypes were visible in the liquid-grown cultures. Here, we observed rhizoid formation as well as brownish inclusions. Indeed, such inclusions also appeared in solid-grown SAG164.80 as well as liquid-grown SAG 650-1—the latter of which is a strain relative of MZCH 240. Despite them being strain relatives, we noticed that the strain MZCH 240 appeared to have a

Table 1 Statistical analysis of maximum quantum yield in *M. scalaris* SAG 164.80 over time. Numbers denote p values obtained through Mann-Whitney U tests

	Solid 2h	Solid 4h	Solid 6h	Solid 8h
Solid 4h	0.1508			
Solid 6h	0.4406	0.7789		
Solid 8h	0.1484	0.726	0.7344	
Solid 24h	0.3828	0.1953	0.5469	0.8332
	Liquid 2h	Liquid 4h	Liquid 6h	Liquid 8h
Liquid 4h	0.875			
Liquid 6h	0.25	0.125		
Liquid 8h	0.09751	0.125	0.125	
Liquid 24h	0.125	0.125	0.625	0.25
	Shift 2h	Shift 4h	Shift 6h	Shift 8h
Shift 4h	0.05469			
Shift 6h	0.01563	0.07813		
Shift 8h	0.02917	0.07593	0.833	
Shift 24h	0.02071	0.03906	0.3615	0.5541

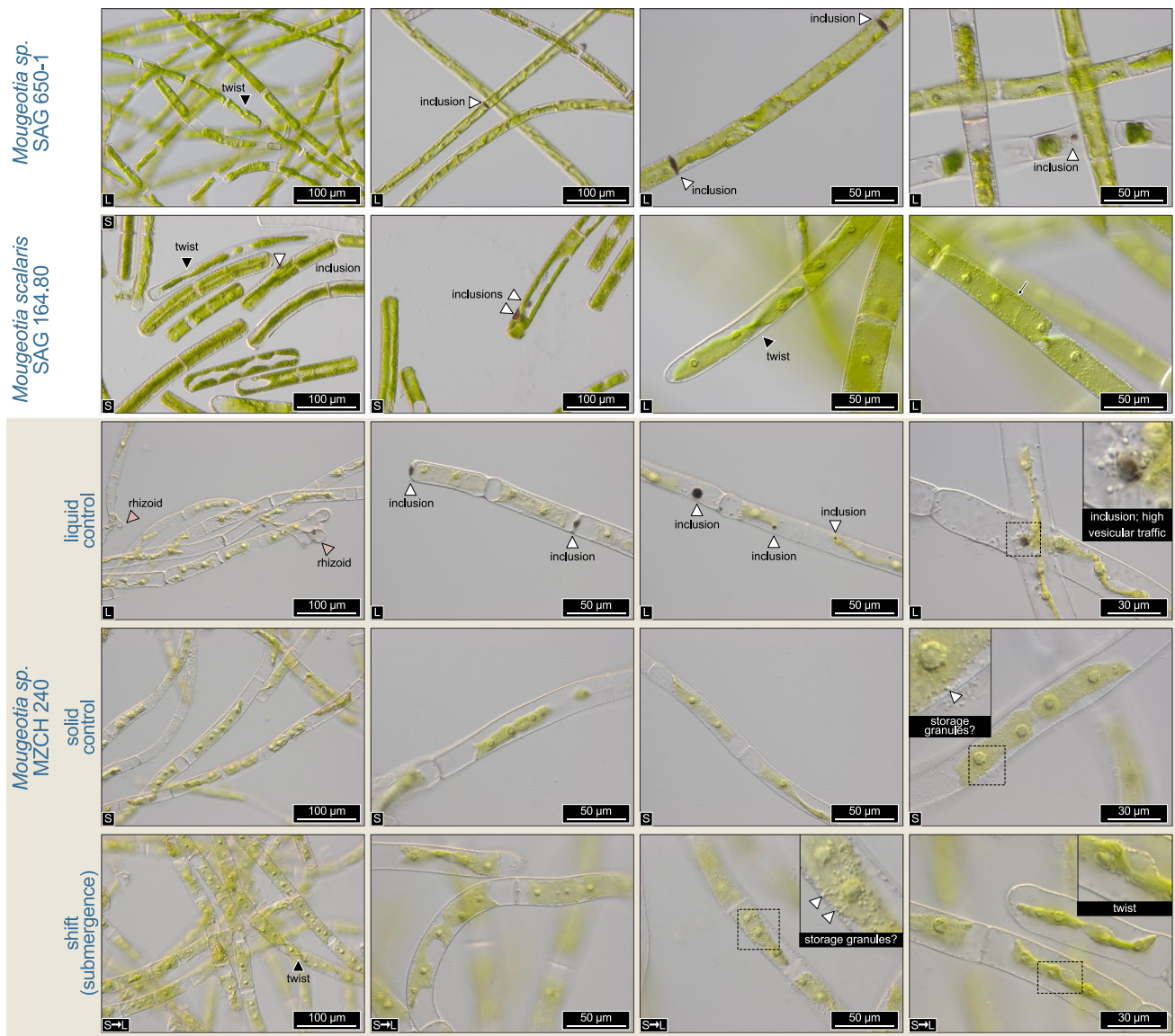


Fig. 3 Notable observations in three *Mougeotia* strains. Nomarski interference contrast micrographs of the strains *Mougeotia* sp. SAG 650-1, *Mougeotia scalaris* SAG 164.80, and *Mougeotia* sp. MZCH 240; the latter was grown in liquid medium, on solid agar plates, and on agar plates and subjected to 24h of submergence in liquid medium (“shift”). The two SAG strains 650-1 and 164.80 were grown either in liquid or on solid MiEB12 Medium, as indicated by the “L” (liquid medium) or “S” (solid medium) on the bottom left

side of the pictures. Notable phenotypic observations include: (a) darkly colored inclusions (sometimes co-occurring with high density of intracellular bodies being trafficked); (b) rhizoid formation in liquid culture; (c) formation of granules, possibly for storage. Also note the twisting chloroplasts, including “edge-on” orientations as a sign for functional chloroplast movement induced by microscope illumination. Labels in the bottom left corner denote: L=liquid-grown, S=solid-grown (agar), S→L=solid-grown and submerged for 24h

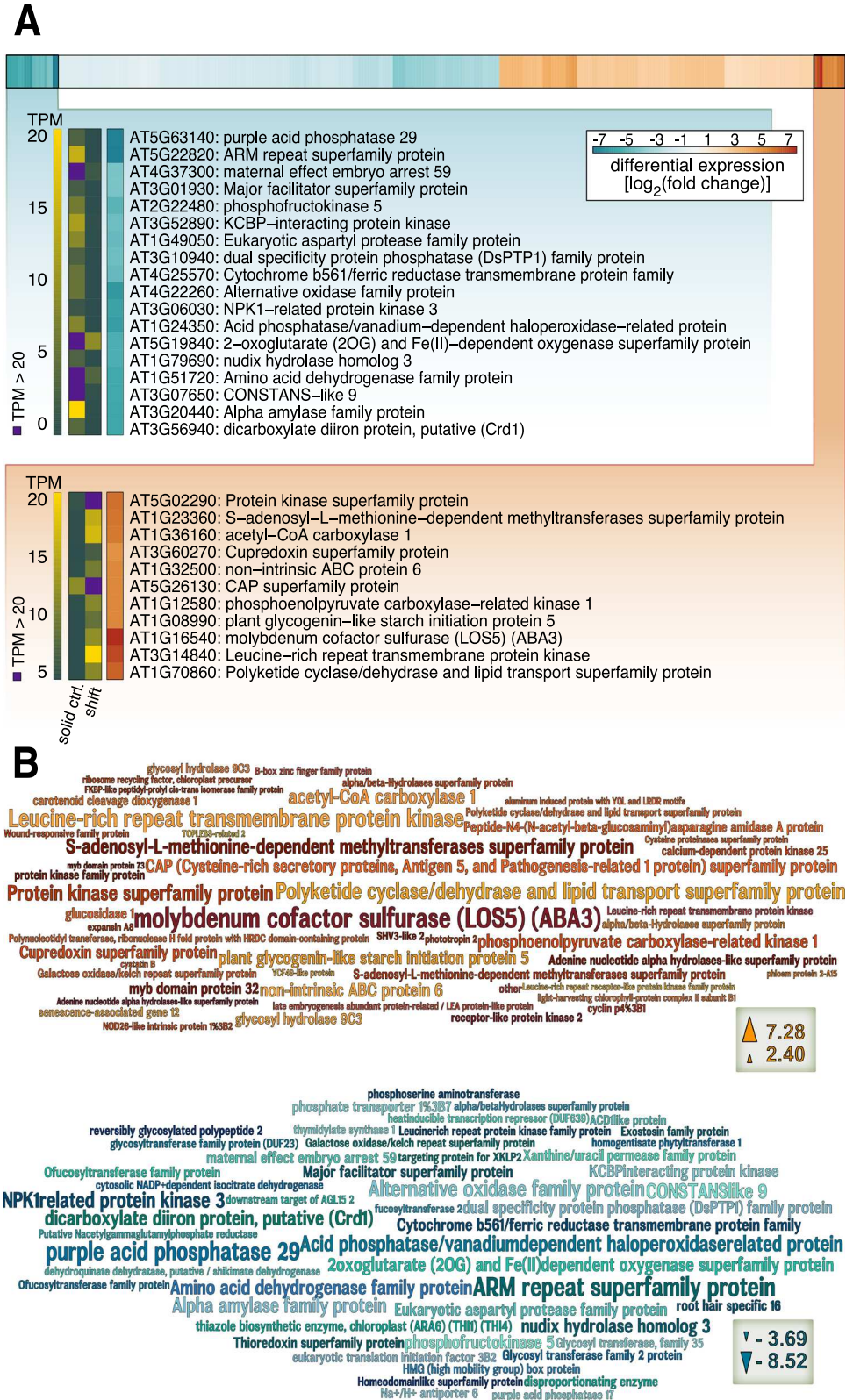
lighter chlorophyllous hue than SAG 650-1, which is however consistent with our previous experience in culturing MZCH 240 (see de Vries et al. 2020).

Together with the gene expression responses, the photophysiological data highlight the fact that the photosynthetic machinery of *Mougeotia* responds to the submergence of the algal filaments in liquid medium. We hence next explored which specific genes might be the key players among these changes.

Responsiveness of genes for light-harvesting components, pigment biosynthesis, and starch metabolism following submergence of *Mougeotia* sp. MZCH 240

To understand which gene expression changes were most pronounced upon submergence, we made use of homology searches against the well-annotated genome of *Arabidopsis thaliana* in combination with the differential transcript

Fig. 4 Top up-/downregulated genes in *Mougeotia* sp. MZCH 240 cultured on solid medium and submerged for 4h versus control on solid medium. **A** A heatmap with all up- (red) or downregulated (blue) genes in *Mougeotia* sp. MZCH 240 based on edgeR analysis of the RNAseq data. Only genes with a significant (Benjamini-Hochberg corrected $p < 0.001$) differential change in gene expression of 2-fold (all differential data are shown as $\log_2[\text{fold change submergence/control}]$, calculated using edgeR) were considered. Using the R package *heatmap* the data were sorted and \log_2 values of clusters of genes with the highest/lowest differential gene expression values are shown. The names and descriptions of corresponding *Arabidopsis thaliana* gene orthologs [prediction based on the reciprocal best BLAST hit (RBBH)] are displayed as well as the corresponding TPM (Transcript per million) values which are shown in a different color gradient (green to yellow). TPM values > 20 are colored in purple; shift = submergence, ctrl. = control. **B** Word clouds of the top 50 up- (red and orange colors) and top 50 downregulated (blue colors) genes in *Mougeotia* sp. MZCH 240 generated with Wordle and based on $\log_2(\text{fold change submergence/control})$, calculated with edgeR. The words represent the names and/or description of Arabidopsis orthologs (prediction based on the RBBH) and the word size corresponds to the differential gene expression change



abundance elicited by submergence of *Mougeotia* sp. MZCH 240. For differential gene expression analyses, we considered only genes that had a Benjamini-Hochberg corrected

$p < 0.001$ and a differential gene expression change of at least 2-fold (Figure 4). Overall, using these criteria, submergence triggered the upregulation of 120 genes (Table 2)

Table 2 120 transcripts that significantly increased in abundance upon submergence in *Mougeotia* sp. MZCH 240

Mousp ID	Best <i>A.t. hit</i>	Annotation	log ₂ (FC)	FDR
Mousp14158_c0_g1_i8	AT1G16540	molybdenum cofactor sulfurase (LOS5) (ABA3)	7.27520966	8.0067E-24
Mousp17078_c0_g3_i2	AT3G14840	Leucine-rich repeat transmembrane protein kinase	6.50770827	3.1038E-22
Mousp17366_c0_g1_i1	AT1G70860	polyketide cyclase/dehydrase and lipid transport	5.97865925	1.8507E-17
Mousp12113_c0_g1_i4	AT5G02290	protein kinase superfamily protein	5.60368112	9.185E-29
Mousp17366_c0_g3_i1	AT1G23360	S-adenosyl-L-methionine-dependent methyltransferases	5.40238661	9.7774E-09
Mousp17745_c1_g1_i11	AT1G36160	acetyl-CoA carboxylase 1	5.37604779	4.3006E-25
Mousp17215_c0_g5_i2	AT1G32500	non-intrinsic ABC protein 6	4.89640786	3.5159E-10
Mousp13170_c0_g1_i1	AT1G12580	phosphoenolpyruvate carboxylase-related kinase 1	4.84629602	8.928E-17
Mousp17366_c0_g2_i3	AT1G08990	plant glycogenin-like starch initiation protein 5	4.80921473	1.6156E-11
Mousp15175_c0_g2_i6	AT5G26130	Cysteine-rich secretory, Antigen 5, and Pathogenesis-related 1	4.76084995	9.7774E-09
Mousp14442_c0_g1_i1	AT3G60270	Cupredoxin superfamily protein	4.52531814	1.3694E-09
Mousp15384_c0_g2_i2	AT4G34990	myb domain protein 32	4.09225758	4.8567E-33
Mousp17772_c0_g1_i14	AT4G11050	glycosyl hydrolase 9C3	3.99810327	2.8402E-10
Mousp16800_c0_g1_i6	AT1G67490	glucosidase 1	3.81940181	1.062E-13
Mousp17501_c0_g1_i5	AT3G14920	Peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase A	3.7028685	1.4233E-09
Mousp17241_c0_g1_i6	AT1G16650	S-adenosyl-L-methionine-dependent methyltransferases	3.54376221	2.6424E-07
Mousp16885_c0_g1_i12	AT3G58450	Adenine nucleotide alpha hydrolases-like	3.43211036	3.7452E-12
Mousp14398_c0_g1_i5	AT3G63520	carotenoid cleavage dioxygenase 1	3.40138953	0.00008287
Mousp17219_c4_g1_i2	AT5G26150	protein kinase family protein	3.27932949	2.8364E-10
Mousp12560_c0_g1_i2	AT2G35890	calcium-dependent protein kinase 25	3.2585986	1.3769E-06
Mousp13723_c0_g1_i1	AT5G45890	senescence-associated gene 12	3.24606174	2.5541E-09
Mousp17422_c0_g1_i1	AT4G11050	glycosyl hydrolase 9C3	3.22808696	0.000010803
Mousp17708_c0_g2_i11	AT3G02130	receptor-like protein kinase 2	3.18914671	0.000023637
Mousp15137_c0_g1_i4	AT2G42450	alpha/beta-Hydrolases superfamily protein	3.10782404	4.3225E-13
Mousp13988_c0_g1_i5	none	none	3.10330699	6.974E-15
Mousp16061_c0_g3_i2	AT1G16250	Galactose oxidase/kelch repeat	3.06024435	1.1042E-16
Mousp16214_c0_g1_i1	AT2G44740	cyclin	3.05199502	2.5095E-15
Mousp14673_c1_g1_i14	AT1G66970	SHV3-like 2	2.98725544	2.3537E-07
Mousp17516_c0_g3_i2	AT4G20140	Leucine-rich repeat transmembrane protein kinase	2.93958975	2.9611E-09
Mousp17666_c0_g1_i1	AT5G20520	alpha/beta-Hydrolases superfamily protein	2.89768053	1.0882E-06
Mousp16146_c1_g1_i34	AT1G55960	Polyketide cyclase/dehydrase and lipid transport	2.89145041	5.3538E-08
Mousp15641_c0_g6_i2	AT2G32415	Polynucleotidyl transferase, ribonuclease H with HRDC domain	2.78988514	9.3814E-07
Mousp15363_c0_g1_i15	AT5G58140	phototropin 2	2.75782905	2.2375E-07
Mousp16839_c0_g1_i4	AT1G19660	Wound-responsive family protein	2.75208096	0.000021332
Mousp16811_c0_g3_i2	AT2G40610	expansin A8	2.73396604	0.00082134
Mousp13223_c0_g1_i6	AT4G18910	NOD26-like intrinsic protein 1%3B2	2.7251728	1.2886E-10
Mousp15049_c1_g5_i1	AT2G21320	B-box zinc finger family protein	2.68212139	7.8998E-14
Mousp15748_c2_g2_i3	AT3G19430	late embryogenesis abundant	2.66385694	3.9849E-08
Mousp16895_c0_g1_i9	AT3G19400	Cysteine proteinases superfamily protein	2.52775348	0.00051691
Mousp17457_c0_g2_i6	AT3G63190	ribosome recycling factor, chloroplast precursor	2.52623134	3.4766E-12
Mousp16770_c0_g2_i2	AT2G34430	light-harvesting chlorophyll-protein complex II subunit B1	2.52608987	0.00040624
Mousp17814_c0_g1_i1	AT3G12490	cystatin B	2.51128417	3.9795E-09
Mousp15345_c0_g4_i4	AT3G22850	aluminum induced protein with YGL and LRDR motifs	2.49160672	8.5789E-07
Mousp16831_c0_g1_i5	AT4G08850	Leucine-rich repeat receptor-like protein kinase	2.47066699	2.3962E-07
Mousp12430_c0_g1_i4	AT4G37260	myb domain protein 73	2.44626078	2.1144E-07
Mousp17048_c2_g5_i2	AT4G22830	YCF49-like protein	2.43736515	1.4633E-09
Mousp13966_c0_g1_i1	AT3G16830	TOPLLESS-related 2	2.42899009	2.509E-07
Mousp12564_c0_g1_i1	AT3G53000	phloem protein 2-A15	2.4230952	3.8864E-10
Mousp15097_c0_g1_i4	AT2G43560	FKBP-like peptidyl-prolyl cis-trans isomerase	2.42256699	1.2101E-10
Mousp15383_c0_g1_i2	AT1G09740	Adenine nucleotide alpha hydrolases-like	2.39874357	0.00028434

Table 2 (continued)

Mousp ID	Best <i>A.t. hit</i>	Annotation	log ₂ (FC)	FDR
Mousp17870_c0_g2_i4	AT2G46580	Pyridoxamine 5'-phosphate oxidase	2.39561141	1.7197E-12
Mousp15748_c2_g3_i4	AT3G19430	late embryogenesis abundant	2.39052414	5.2306E-08
Mousp17536_c0_g2_i15	AT4G16760	acyl-CoA oxidase 1	2.37665903	0.000010184
Mousp17563_c0_g2_i4	AT1G13980	sec7 domain-containing protein	2.36154393	0.00012739
Mousp17005_c2_g1_i2	AT5G54370	late embryogenesis abundant	2.34756478	0.000050599
Mousp17009_c0_g2_i11	AT4G33010	glycine decarboxylase P-protein 1	2.34716303	0.000027572
Mousp13949_c0_g1_i1	AT5G19360	calcium-dependent protein kinase 34	2.3443775	0.00023592
Mousp14435_c0_g1_i1	AT3G22750	Protein kinase superfamily protein	2.29061994	1.2152E-11
Mousp16006_c1_g2_i6	AT2G15010	Plant thionin	2.28071231	0.000001202
Mousp17583_c1_g1_i5	AT1G08550	non-photochemical quenching 1	2.26244193	6.0714E-06
Mousp17901_c2_g2_i8	AT5G14580	polyribonucleotide nucleotidyltransferase	2.24801086	0.00008991
Mousp15753_c0_g1_i21	AT2G34260	transducin family protein / WD-40 repeat	2.24643147	7.4559E-07
Mousp15748_c2_g4_i7	AT3G19430	late embryogenesis abundant	2.24637151	6.5626E-06
Mousp15175_c0_g1_i6	AT2G14610	pathogenesis-related protein 1	2.24594398	0.000030745
Mousp17901_c2_g3_i1	none	none	2.24550218	0.000023275
Mousp16876_c0_g5_i2	AT3G52140	tetratricopeptide repeat (TPR)-containing protein	2.22759061	0.000030632
Mousp17754_c1_g2_i1	AT5G41460	transferring glycosyl group transferase (DUF604)	2.22550195	1.2101E-10
Mousp10496_c0_g1_i1	AT4G33880	ROOT HAIR DEFECTIVE 6-LIKE 2	2.21640294	7.7321E-10
Mousp14422_c0_g1_i6	AT1G14870	PLANT CADMIUM RESISTANCE 2	2.21344381	8.4799E-06
Mousp13841_c0_g1_i3	AT2G24440	selenium binding protein	2.1868762	1.7433E-10
Mousp17685_c0_g1_i2	AT4G00260	Transcriptional factor B3 family protein	2.173197	0.00032902
Mousp17103_c0_g2_i3	AT2G37560	origin recognition complex second largest subunit 2	2.15944614	0.000088228
Mousp14784_c2_g1_i6	AT2G21940	shikimate kinase 1	2.14189653	4.7278E-09
Mousp16295_c0_g1_i5	AT1G31420	Leucine-rich repeat protein kinase	2.14136017	0.00020268
Mousp17228_c0_g3_i14	AT2G25185	Defensin-like (DEFL) family protein	2.13194267	0.00075013
Mousp14776_c0_g1_i2	AT5G15330	SPX domain-containing protein 4	2.12256734	3.8238E-10
Mousp15459_c0_g2_i1	AT1G44575	Chlorophyll A-B binding family protein	2.11945161	1.1268E-07
Mousp17556_c0_g1_i6	AT5G64290	dicarboxylate transport 2.1	2.11755201	0.00092571
Mousp16715_c1_g1_i6	AT2G33855	transmembrane protein	2.06354721	7.3282E-09
Mousp12292_c0_g1_i2	AT5G09650	pyrophosphorylase 6	2.02084136	1.0549E-06
Mousp15459_c0_g3_i1	AT1G44575	Chlorophyll A-B binding family protein	2.01519992	2.6424E-07
Mousp11032_c0_g1_i1	AT2G36930	zinc finger (C2H2 type) family protein	1.99550424	3.5079E-07
Mousp15459_c1_g1_i1	AT1G44575	Chlorophyll A-B binding family protein	1.96857574	1.5821E-06
Mousp11772_c0_g1_i3	AT1G22170	Phosphoglycerate mutase family protein	1.96630859	1.2869E-06
Mousp17393_c0_g3_i1	AT2G40490	Uroporphyrinogen decarboxylase	1.94764795	0.000013015
Mousp15227_c0_g1_i2	AT5G65230	myb domain protein 53	1.94157129	1.3559E-06
Mousp16477_c0_g4_i7	AT5G52975	egg cell-secreted-like protein (DUF1278)	1.92832108	0.00010985
Mousp15882_c0_g1_i1	AT2G19540	Transducin family protein / WD-40 repeat family protein	1.92449346	0.00096677
Mousp16664_c0_g3_i1	AT3G12410	Polynucleotidyl transferase, ribonuclease H-like	1.91165287	0.00028833
Mousp16466_c0_g1_i3	AT2G35120	Single hybrid motif superfamily protein	1.82700212	1.5271E-06
Mousp15769_c0_g1_i1	AT4G24230	acyl-CoA-binding domain 3	1.8206261	0.00013362
Mousp16717_c0_g1_i10	AT3G19430	late embryogenesis abundant	1.82053718	0.00022751
Mousp17443_c0_g1_i9	AT4G35000	ascorbate peroxidase 3	1.81353441	2.5071E-07
Mousp12426_c0_g1_i4	AT5G22140	FAD/NAD(P)-binding oxidoreductase family protein	1.79593008	0.00026667
Mousp15265_c0_g1_i8	AT5G02160	transmembrane protein	1.73674051	0.00003623
Mousp14642_c0_g1_i3	AT4G15520	tRNA/rRNA methyltransferase (SpoU) family protein	1.73401135	0.00087276
Mousp12053_c0_g1_i1	AT5G49300	GATA transcription factor 16	1.70744995	0.000012895
Mousp17024_c0_g1_i31	AT1G29900	carbamoyl phosphate synthetase B	1.69430823	0.00029494
Mousp14546_c0_g1_i2	AT5G48300	ADP glucose pyrophosphorylase 1	1.67256288	0.000055648
Mousp16932_c4_g2_i3	AT1G78430	ROP interactive partner 2	1.67009459	0.00026667

Table 2 (continued)

Mousp ID	Best <i>A.t. hit</i>	Annotation	log ₂ (FC)	FDR
Mousp17693_c0_g2_i9	AT5G04270	DHHC-type zinc finger family protein	1.6547086	0.000030487
Mousp15496_c2_g7_i4	AT1G20140	SKP1-like 4	1.63691483	0.00012795
Mousp14376_c0_g1_i13	AT3G21150	B-box 32	1.62615568	0.00022751
Mousp16045_c0_g1_i2	AT5G13680	IKI3 family protein	1.61640969	0.000031256
Mousp17530_c2_g2_i38	AT3G63380	ATPase E1-E2 / haloacid dehalogenase-like hydrolase	1.61216412	0.00086341
Mousp13515_c0_g1_i2	AT5G12180	calcium-dependent protein kinase 17	1.60553028	0.000012596
Mousp17689_c0_g1_i7	AT4G30990	ARM repeat superfamily protein	1.59876228	0.00092316
Mousp10275_c0_g1_i1	AT4G14890	2Fe-2S ferredoxin-like superfamily protein	1.58967395	0.0003602
Mousp17110_c0_g2_i14	AT3G05060	NOP56-like pre RNA processing ribonucleoprotein	1.51185847	0.00051743
Mousp16289_c1_g3_i6	AT3G45190	SIT4 phosphatase-associated family protein	1.51073535	0.00052452
Mousp13265_c0_g1_i2	AT1G13580	LAG1 longevity assurance-like protein	1.47973166	0.000049111
Mousp17103_c0_g1_i2	AT5G48630	Cyclin family protein	1.45228723	0.00043854
Mousp15997_c0_g1_i6	AT5G37850	pfkB-like carbohydrate kinase family protein	1.45112666	0.00029494
Mousp15166_c1_g3_i2	AT4G27600	pfkB-like carbohydrate kinase family protein	1.44542141	0.00037996
Mousp17238_c0_g3_i3	AT3G43520	Transmembrane proteins 14C	1.36066059	0.00067169
Mousp16518_c0_g3_i2	AT4G23890	NAD(P)H-quinone oxidoreductase subunit S	1.35071287	0.00040989
Mousp17358_c1_g1_i9	AT3G04460	peroxin-12	1.3027287	0.00081849
Mousp17756_c0_g1_i13	AT4G22890	PGR5-LIKE A	1.30095184	0.0007896
Mousp13036_c1_g1_i1	AT4G29350	profilin 2	1.2830376	0.00092482
Mousp17102_c1_g1_i1	AT3G61070	peroxin 11E	1.26175474	0.00093898

Differential changes in transcript abundance (*FC*, fold change) in the samples taken 4h after submergence in liquid medium were calculated versus solid control and log₂-transformed using edgeR; *FDR* false discovery rate denotes Benjamini-Hochberg-corrected *p* values; *A.t. Arabidopsis thaliana*

and the downregulation of 171 genes (Supplementary Figure S2). Again, photosynthesis-related gene expression patterns stood out—both concerning genes relevant to the light reaction and those of downstream processes, such as three genes putatively coding for chlorophyll *a/b*-binding proteins (4.3-fold, 4.0-fold, and 3.9-fold upregulation) and a gene putatively encoding a light-harvesting component showed induction (Figure S1; 5.8-fold upregulation).

A *Mougeotia* sp. transcript homologous to *AtABA3* corresponded to the highest gene expression change (i.e. differential change in transcript abundance); it was up-regulated 154.9-fold following the shift from dry to wet. *ABA3* codes for a cytosolic molybdenum cofactor sulfurase that converts the carotenoid-derived abscisic aldehyde into the phytohormone abscisic acid (ABA). Despite the fact that several Zygnematophyceae have genes for the ABA receptors (de Vries et al. 2018; Cheng et al. 2019), these likely act in an ABA-independent function (Sun et al. 2019). We interpret the induction of the *ABA3* homolog rather as a readout of the aforementioned regulation of pigments (in this case, carotenoid metabolism) and photosynthesis-associated genes expression patterns that impact overall plastid physiology. In line with this, we also found regulation of violaxanthin de-epoxidase (4.8-fold upregulation) and a carotenoid cleavage dioxygenase (a homolog of CCD1; 10.6-fold upregulation).

Carotenoid cleavage-derived metabolites are well known signaling molecules in plant cells—especially elicited upon environmental cues (Hou et al. 2016). Indeed, heat-induced changes in the expression of CCDs were observed for *Mougeotia* sp. (de Vries et al. 2020). Another aspect that needs to be taken into consideration is the adjustment of pigment profiles upon acclimating to a changing habitat; in an aquatic environment, not only the intensity but also the quality of light differs. Here, *Mougeotia* is a system rich in experimental history: in this algal genus, extensive work on chloroplast movement dependent on light qualities sensed by photoreceptors were carried out (Wagner and Klein 1981). Interestingly, Zygnematophyceae such as *Mougeotia* stand out by having chimeric photoreceptors containing domains of the red light phytochromes and blue light phototropins, the so-called neochromes (in our assembly *Mousp17450_c0_g1*; Data S1; (Suetsuga et al. 2006; Li et al. 2015). Responses regulated by these photoreceptors include chloroplast movement (note some of the twisting chloroplasts in Figure 3). We did not find clear signs for the differential regulation of genes related to light quality signaling (e.g. non-significant 2-fold downregulation of the phytochrome B homolog *Mousp17540_c0_g1*); further, the neochrome transcript *Mousp17450_c0_g1* was induced upon submergence, with an average TPM of 0.15 in solid control and 0.55 upon

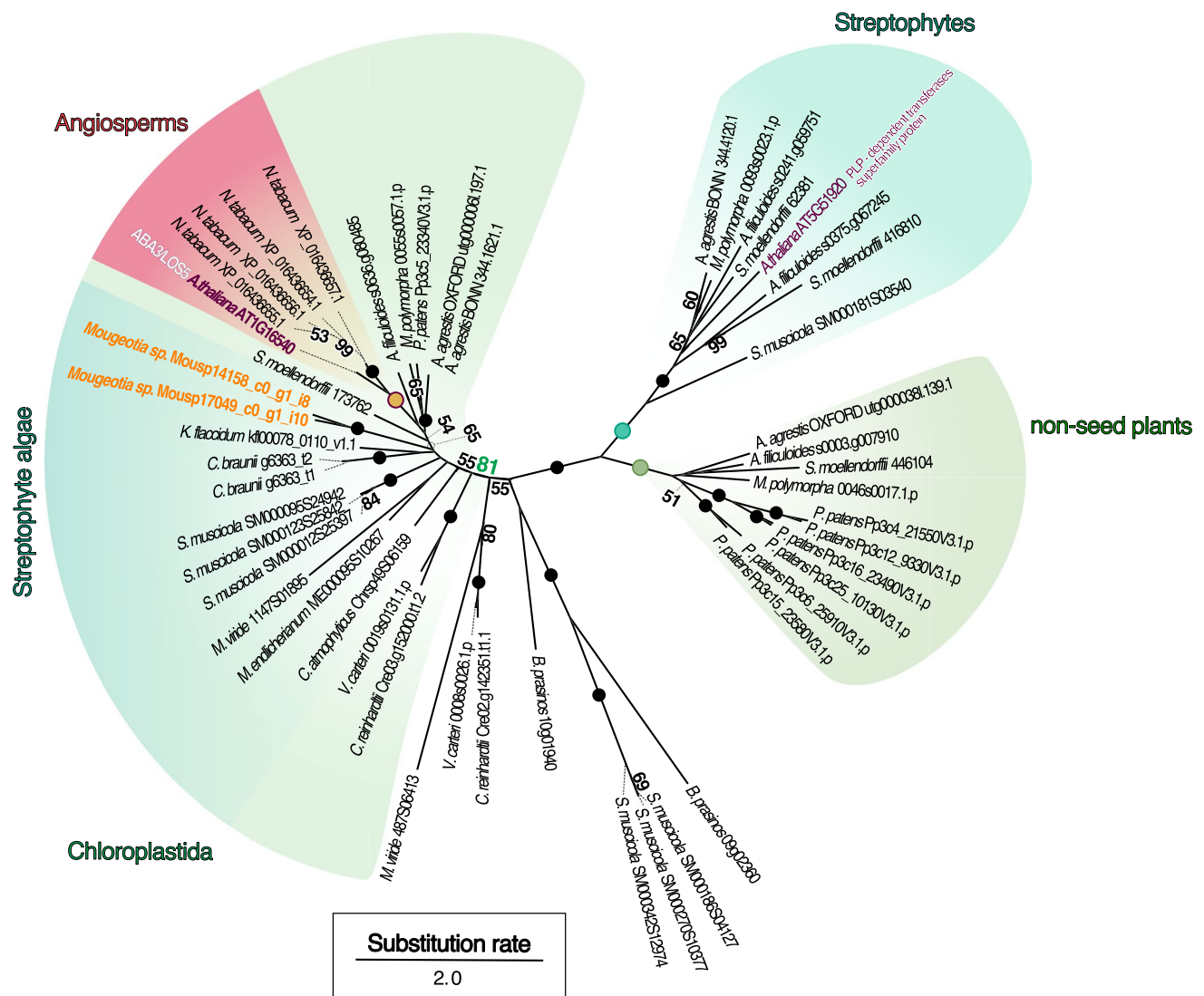


Fig. 5 Phylogenetic framework for the putative ABA3 sequences identified in *Mougeotia* sp. MZCH 240. Phylogeny of homologs for the molybdenum cofactor sulfuryase ABA3. Two homologs of ABA3 (Mousp14158_c0_g1_i8, Mousp17049_c0_g1_i10), the first of which was the most up-regulated gene in *Mougeotia* sp. MZCH 240 upon submergence, were aligned with 48 ABA3 homologs

detected in diverse land plants, streptophyte algae, and chlorophyte algae. Homologs were aligned and an unrooted maximum-likelihood phylogeny was computed using WAG+F+I+G4 (chosen according to BIC) as model for protein evolution and 100 bootstrap replicates. Bootstrap values <50 are not shown in the figure; maximum bootstrap support is indicated by a filled dot

4h submergence—however, as the numbers give away, it was expressed at such a low level that it was excluded from the analyses (see Material and Methods). Overall, it is conceivable that sensing the different spectral qualities of light when shifting to submergence is important and deserves further investigation.

To explore whether the *Mougeotia* sp. ABA3 homolog we detected is likely an ABA3 ortholog, we performed a phylogenetic analysis. We used BLASTp to mine a phylodiverse protein dataset for ABA3 homologs, MAFFT (Katoh and Standley 2013) to align all putative ABA3 sequences, and IQ-TREE (v1.5.5; Nguyen et al. 2015) to

construct a maximum likelihood phylogeny (Figure 5). The putative ABA3 homolog detected in *Mougeotia* sp. (Mousp14158_c0_g1_i8) fell, together with a potential paralog (Mousp17049_c0_g1_i10), into a moderately supported (65% bootstrap value) clade of land plant sequences. This clade was, however, nested in a more highly supported (81% bootstrap) clade of putative molybdenum cofactor sulfurases from across Chloroplastida. Thus, the ABA3 homolog detected in *Mougeotia* sp. seems to fall into the orthogroup of ABA3-type Molybdenum cofactor sulfurases that is conserved across Chloroplastida.

Green algae and land plants store photosynthate as starch. The buildup of starch appears to depend on the action of PLANT GLYCOGENIN-LIKE STARCH INITIATION PROTEINs (PGSIP; Chatterjee et al. 2005). Interestingly, we found a homolog of *PGSIP5* (AT1G08990) that is strongly induced (28.0-fold up) upon submergence. In light of the changes to the photosynthesis machinery, it is logical to also find genes associated with the downstream buildup of water-insoluble starch; the buildup of reserves appears a common theme among filamentous Zygnematophyceae that are challenged with environmental fluctuations (Pichrtová et al. 2016; Arc et al. 2020; de Vries and Ischebeck 2020). Indeed, the only enriched GO-term process was among the downregulated genes; there, we found that the GO-term “cellular carbohydrate catabolic process” (GO:0044275; *p* value 6.71x10⁻⁴) was enriched.

A homolog of a gene encoding a purple acid phosphatase (PAP) was found as the second most downregulated *Mougeotia* sp. gene (Mousp11308_c0_g1_i1; 301.0-fold downregulated); the resulting *Mougeotia* sp. protein bears a signal peptide (likelihood of 0.99 on TargetP-2.0), thus resembling the repertoire of secreted land plant PAPs with diverse functions in response to shifts in environmental conditions and nutrient availability (Bozzo et al. 2002; Kaida et al. 2010; Wang et al. 2011). It is noteworthy that, in a phylogenetic analysis, the *Mougeotia* sp. PAP fell into a clade of chlorophyte and streptophyte green algae, which formed a monophylum distinct from land plant PAPs (Figure S3).

Finally, we found differentially expressed *Mougeotia* sp. genes that are classically associated with pathogen response, including a gene putatively encoding a leucine-rich repeat transmembrane protein, (homologs of AT3G14840 and AT4G20140 were 91.0 and 7.7-fold up, respectively) and CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein; AT5G26130; 27.1-fold up). Such proteins are, however, equally often a sign of stress elicited by various changes in the environment (Creff et al. 2019 (AT4G20140); Le et al. 2014 (AT3G14840), Chien et al. 2015 (AT5G26130))—they might simply be a read-out of the interwoven network that underpins environmental sensing. In line with this, a gene homologous to protein kinase-encoding *AT5G02290* showed clear induction (48.6-fold upregulation); this kinase might be involved in various signaling processes and speaks to the response of *Mougeotia* sp. to the changing environment. Indeed, several genes that speak to a general stress response were up-regulated. These included five LATE EMBRYOGENESIS ABUNDANT (LEA) homologs (6.3-fold, 5.2-fold, 5.1-fold, 4.7-fold, and 3.5-fold up-regulated), which are classical factors responsive to various abiotic stressors in other systems (Ingram and Bartels 1996; Hundertmark and Hinch 2008).

Conclusion

We observed that submergence of *Mougeotia* triggered a conspicuous set of differentially regulated genes associated with changes in several photosynthesis and primary carbon metabolic pathways, suggesting remodeling of the photosystem apparatuses. This notion is supported by the observation that (a) various other photosynthesis-associated genes changed their expression and (b) slight but significant changes in the photochemical performance measured through the maximum quantum yield (F_v/F_m) were observed. Additionally, genes that speak to a remodeling of the pigment composition were regulated. It is conceivable that the composition of accessory pigments is being adjusted in response to the altered quality of light triggered by submergence. Altogether, our data suggest that some of the foremost adjustments that these filamentous zygnematophycean algae undergo during dry-to-wet transition are related to photophysiological acclimation; an assessment of the degree to which this holds true in the ecophysiological setting of temporary freshwater bodies is bound to be illuminating.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00709-021-01730-1>.

Acknowledgements We thank Prof. Dr. Christiane Gatz and Dr. Guido Kriete for giving us access to the ImagMAX/L PAM in the Department of Plant Molecular Biology and Physiology. J.M.R.F.-J. is grateful for being supported by the Ph.D. program “Microbiology and Biochemistry” within the framework of the “Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences” (GGNB) at the University of Goettingen.

Funding Open Access funding enabled and organized by Projekt DEAL. Work in the lab of JdV is supported by funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 852725; ERC Starting Grant “TerreStrIAL”). Research in the lab of JMA was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2014-05871). J.d.V. received grants from the German Research Foundation (DFG) within the framework of the Priority Programme “MAdLand – Molecular Adaptation to Land: Plant Evolution to Change” (SPP 2237; VR 132/4-1), in which J.M.R.F.-J partakes as associate member. K.v.S. received grants from Deutsche Forschungsgemeinschaft (Schw687/13-1).

Data availability All data generated or analyzed during this study are included in this published article (and its [supplementary information](#) files), and the public databases of the NCBI: all RNAseq read data have been uploaded to the NCBI SRA. The reads from the control samples are available under the run IDs SRR9083693, SRR9083694, SRR9083695, SRR9083697, SRR9083698, SRR9083699; liquid treatment is available under the run IDs SRR9083681, SRR9083682, SRR9083688 (<https://www.ncbi.nlm.nih.gov/sra?term=SRP198800>). The reference assembly is publicly available under NCBI BioProject PRJNA543475 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA543475>).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arc E, Pichrtová M, Kranner I, Holzinger A (2020) Pre-akinete formation in *Zygnema* sp. from polar habitats is associated with metabolite re-arrangement. *J Exp Bot* 57:289–289. <https://doi.org/10.1093/jxb/eraa123>
- Banks JA, Nishiyama T, Hasebe M et al (2011) The selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332:960–963. <https://doi.org/10.1126/science.1203810>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bowman JL, Kohchi T, Yamato KT et al (2017) Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* 171:287–299.e15. <https://doi.org/10.1016/j.cell.2017.09.030>
- Bozzo GG, Raghothama KG, Plaxton WC (2002) Purification and characterization of two secreted purple acid phosphatase isozymes from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures: secreted acid phosphatases of P_i-starved tomato cells. *Eur J Biochem* 269:6278–6286. <https://doi.org/10.1046/j.1432-1033.2002.03347.x>
- Chatterjee Manash, Berbezy P, Vyas D, Coates S, Barsby T (2005) Reduced expression of a protein homologous to glycogenin leads to reduction of starch content in *Arabidopsis* leaves. *Plant Science* 168(2):501–509. <https://doi.org/10.1016/j.plantsci.2004.09.015>
- Cheng S, Xian W, Fu Y et al (2019) Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* 179:1057–1067.e14. <https://doi.org/10.1016/j.cell.2019.10.019>
- Chien PS, Nam HG, Chen YR (2015) A salt-regulated peptide derived from the CAP superfamily protein negatively regulates salt-stress tolerance in *Arabidopsis*. *J Exp Bot* 66:5301–5313. <https://doi.org/10.1093/jxb/erv263>
- Christa G, Cruz S, Jahns P et al (2017) Photoprotection in a monophyletic branch of chlorophyte algae is independent of energy-dependent quenching (qE). *New Phytol* 214:1132–1144. <https://doi.org/10.1111/nph.14435>
- Correa-Galvis V, Poschmann G, Melzer M et al (2016) PsbS interactions involved in the activation of energy dissipation in *Arabidopsis*. *Nat Plants* 2:15225. <https://doi.org/10.1038/nplants.2015.225>
- Creff A, Brocard L, Joubès J, Taconnat L, Doll NM, Marsollier AC, Ingram G et al (2019) A stress-response-related inter-compartmental signalling pathway regulates embryonic cuticle integrity in *Arabidopsis*. *PLoS Genet* 15:1–28. <https://doi.org/10.1371/journal.pgen.1007847>
- de Vries J, Ischebeck T (2020) Ties between stress and lipid droplets pre-date seeds. *Trends Plant Sci* 25:1203–1214. <https://doi.org/10.1016/j.tplants.2020.07.017>
- de Vries PJR, Simons J, van Beem AP (1983) Sporopollenin in the spore wall of *Spirogyra* (Zygnemataceae, Chlorophyceae). *Acta Botanica Neerlandica* 32:25–28. <https://doi.org/10.1111/j.1438-8677.1983.tb01674.x>
- de Vries J, de Vries S, Slamovits CH et al (2017) How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant Cell Physiol* 58:934–945. <https://doi.org/10.1093/pcp/pcx037>
- de Vries J, Curtis BA, Gould SB, Archibald JM (2018) Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc Natl Acad Sci U S A* 115:E3471–E3480. <https://doi.org/10.1073/pnas.1719230115>
- de Vries J, Vries S, Curtis BA et al (2020) Heat stress response in the closest algal relatives of land plants reveals conserved stress signaling circuits. *Plant J* 324:1064–1024. <https://doi.org/10.1111/tjp.14782>
- de Vries S, Fürst-Jansen JMR, Irisarri I, Dhabalia Ashok A, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J (2021) The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *Plant J* 107:975–1002. <https://doi.org/10.1111/tjp.15387>
- Eden E, Navon R, Steinfeld I et al (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform* 10:48. <https://doi.org/10.1186/1471-2105-10-48>
- FASTQC (2018) A quality control tool for high throughput sequence data. Available at www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed September 15, 2018.
- Foyer CH, Lelandais M, Kunert KJ (1994) Photooxidative stress in plants. *Physiol Plant* 92:696–717. <https://doi.org/10.1111/j.1399-3054.1994.tb03042.x>
- Friedl T, Lorenz M (2012) The Culture Collection of Algae at Göttingen University (SAG): a biological resource for biotechnological and biodiversity research. *Procedia Environ Sci* 15:110–117. <https://doi.org/10.1016/j.proenv.2012.05.015>
- Fürst-Jansen JMR, de Vries S, de Vries J (2020) Evo-physio: on stress responses and the earliest land plants. *J Exp Bot* 66:4–16. <https://doi.org/10.1093/jxb/eraa007>
- Gerotto C, Morosinotto T (2013) Evolution of photoprotection mechanisms upon land colonization: evidence of PSBS-dependent NPQ in late Streptophyte algae. *Physiol Plant* 149:583–598. <https://doi.org/10.1111/ppl.12070>
- Guillard RRL (1975) Culture of phytoplankton for feeding marine invertebrates. In: Smith WL, Chanley MH (eds) Culture of marine invertebrate animals. Plenum Book Publ. Corp, New York, pp 29–60
- Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Herburger K, Holzinger A (2015) Localization and Quantification of Callose in the Streptophyte Green Algae *Zygnema* and *Klebsormidium*: Correlation with Desiccation Tolerance. *Plant Cell Physiol* 56:2259–2270. <https://doi.org/10.1093/pcp/pcv139>
- Holzinger A, Pichrtová M (2016) Abiotic stress tolerance of charophyte green algae: new challenges for omics techniques. *Front Plant Sci* 7:273–217. <https://doi.org/10.3389/fpls.2016.00678>
- Holzinger A, Kaplan F, Blaas K, Zechmann B, Komsic-Buchmann K, Becker B (2014) Transcriptomics of desiccation tolerance in the streptophyte green alga *Klebsormidium* reveal a land plant-like defense reaction. *PLoS One* 9:e110630
- Holzinger A, Albert A, Aigner S, Uhl J, Schmitt-Kopplin P, Trumhová K, Pichrtová M (2018) Arctic, antarctic, and temperate green algae *Zygnema* spp. under UV-B stress: vegetative cells perform

- better than pre-akinetes. *Protoplasma* 255:1239–1252. <https://doi.org/10.1007/s00709-018-1225-1>
- Hori K, Maruyama F, Fujisawa T et al (2014) Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat Commun* 5:3978. <https://doi.org/10.1038/ncomms4978>
- Hou X, Rivers J, León P et al (2016) Synthesis and function of apocarotenoid signals in plants. *Trends Plant Sci* 21:792–803. <https://doi.org/10.1016/j.tplants.2016.06.001>
- Hundertmark M, Hinch DK (2008) LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9:118–122. <https://doi.org/10.1186/1471-2164-9-118>
- Hutin C, Nussaume L, Moise N et al (2003) Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proc Natl Acad Sci U S A* 100:4921–4926. <https://doi.org/10.1073/pnas.0736939100>
- Ingram J, Bartels D (1996) The molecular basis of dehydration tolerance in plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:377–403. <https://doi.org/10.1146/annurev.arplant.47.1.377>
- Irisarri I, Darienko T, Pröschold T, Fürst-Jansen JMR, Jamy M, de Vries J (2021) Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc R Soc B* 288:20212168. <https://doi.org/10.1098/rspb.2021.2168>
- Jahns P, Holzwarth AR (2012) The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *BBA-Bioenergetics* 1817:182–193. <https://doi.org/10.1016/j.bbabi.2011.04.012>
- Jiao C, Sørensen I, Sun X et al (2020) The Penium margaritaceum Genome: Hallmarks of the Origins of Land Plants. *Cell* 181(P1097-1111):E12. <https://doi.org/10.1016/j.cell.2020.04.019>
- Kaida R, Serada S, Norioka N et al (2010) Potential role for purple acid phosphatase in the dephosphorylation of wall proteins in tobacco cells. *Plant Physiol* 153:603–610
- Kalyaanamoorthy S, Minh BQ, Wong TKF et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
- Karsten U, Lembcke S, Schumann R (2007) The effects of ultraviolet radiation on photosynthetic performance, growth and sunscreen compounds in aeroterrestrial biofilm algae isolated from building facades. *Planta* 225:991–1000
- Karsten U, Herburger K, Holzinger A (2014) Dehydration, temperature, and light tolerance in members of the aeroterrestrial green algal genus *Interfilum* (Streptophyta) from biogeographically different temperate soils. *J Phycol* 50:804–816. <https://doi.org/10.1111/jpy.12210>
- Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Kitzing C, Karsten U (2015) Effects of UV radiation on optimum quantum yield and sunscreen contents in members of the genera *Interfilum*, *Klebsormidium*, *Hormidiella* and *Entransia* (Klebsormidiophyceae, Streptophyta). *Eur J Phycol* 50:279–287. <https://doi.org/10.1080/09670262.2015.1031190>
- Krause GH, Veronotte C, Briantais JM (1982) Photoinduced quenching of chlorophyll fluorescence in intact chloroplasts and algae. Resolution into two components. *Biochim Biophys Acta* 679:116–124. [https://doi.org/10.1016/0005-2728\(82\)90262-6](https://doi.org/10.1016/0005-2728(82)90262-6)
- Lamesch P, Berardini TZ, Li D et al (2012) The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res* 40:D1202–D1210. <https://doi.org/10.1093/nar/gkr1090>
- Lang D, Ullrich KK, Murat F et al (2018) The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J* 93:515–533. <https://doi.org/10.1111/tpj.13801>
- Le MH, Cao Y, Zhang X-C, Stacey G (2014) LIK1, A CERK1-Interacting Kinase, Regulates Plant Immune Responses in *Arabidopsis*. *PLoS ONE* 9(7):e102245. <https://doi.org/10.1371/journal.pone.0102245>
- Leebens-Mack JH, Barker MS, Carpenter EJ et al (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685. <https://doi.org/10.1038/s41586-019-1693-2>
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
- Li X-P, Björkman O, Shih C, Grossman AR, Rosenquist M, Jansson S, Niyogi KK (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* 403:391–395. <https://doi.org/10.1038/35000131>
- Li F-W, Melkonian M, Rothfels CJ et al (2015) Phytochrome diversity in green plants and the origin of canonical plant phytochromes. *Nat Commun* 6:7852. <https://doi.org/10.1038/ncomms8852>
- Li F-W, Nishiyama T, Waller M et al (2020) *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants* 6:259–272. <https://doi.org/10.1038/s41477-020-0618-2>
- Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M et al (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 4(7):460–472. <https://doi.org/10.1038/s41477-018-0188-8>
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*:50–60
- Merchant SS, Prochnik SE, Vallon O et al (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318:245–251. <https://doi.org/10.1126/science.1143609>
- Mikhailyuk T, Glaser K, Holzinger A, Karsten U (2015) Biodiversity of Klebsormidium (Streptophyta) from alpine biological soil crusts (Alps, Tyrol, Austria, and Italy). *J Phycol* 51:750–767. <https://doi.org/10.1111/jpy.12316>
- Montané MH, Dreyer S, Triantaphylidès C, Kloppstech K (1997) Early light-inducible proteins during long-term acclimation of barley to photooxidative stress caused by light and cold: High level of accumulation by posttranscriptional regulation. *Planta* 202:293–302. <https://doi.org/10.1007/s004250050131>
- Moreau H, Verhelst B, Couloux A et al (2012) Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol* 13:R74
- Morris JL, Puttick MN, Clark JW et al (2018) The timescale of early land plant evolution. *Proc Natl Acad Sci U S A* 115:E2274–E2283. <https://doi.org/10.1073/pnas.1719588115>
- Müller P, Li XP, Niyogi KK (2001) Non-photochemical quenching. A response to excess light energy. *Plant Physiol* 125:1558–1566. <https://doi.org/10.1186/gb-2012-13-8-r74>
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Nichols HW (1973) Growth media – freshwater. In: Stein JR (ed) *Handbook of Phycological Methods*. Cambridge University Press, London, pp 16–17
- Nishiyama T, Sakayama H, de Vries J et al (2018) The Chara Genome: secondary complexity and implications for plant terrestrialization. *Cell* 174:448–464.e24. <https://doi.org/10.1016/j.cell.2018.06.033>
- Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K (2017) Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci* 22:53–65

- Peers G, Truong TB, Ostendorf E et al (2009) An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* 462:518–521. <https://doi.org/10.1038/nature08587>
- Permann C, Herburger K, Niedermeier M, Felhofer M, Gierlinger N, Holzinger A (2021) Cell wall characteristics during sexual reproduction of *Mougeotia* sp. (Zygnematophyceae) revealed by electron microscopy, glycan microarrays and RAMAN spectroscopy. *Protoplasma* 258(6):1261–1275. <https://doi.org/10.1007/s00709-021-01659-5>
- Pichrtová M, Remias D, Lewis LA, Holzinger A (2013) Changes in phenolic compounds and cellular ultrastructure of arctic and antarctic strains of *Zygnema* (Zygnematophyceae, Streptophyta) after exposure to experimentally enhanced UV to PAR ratio. *Microb Ecol* 65:68–83
- Pichrtová M, Arc E, Stöggel W, Kranner I, Hájek T, Hackl H, Holzinger A (2016) Formation of lipid bodies and changes in fatty acid composition upon pre-akinete formation in Arctic and Antarctic *Zygnema* (Zygnematophyceae, Streptophyta) strains. *FEMS Microbiol Ecol* 92:fiw096–fiw099. <https://doi.org/10.1093/femsec/fiw096>
- Pouličková A, Zizka Z, Hasler P, Benada O (2007) Zygnematalean zygospores: morphological features and use in species identification. *Folia Microbiol* 52:135–145. <https://doi.org/10.1007/BF02932152>
- Prochnik SE, Umen J, Nedelcu AM et al (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329:223–226. <https://doi.org/10.1126/science.1188800>
- Regensdorff M, Deckena M, Stein M, Borchers A, Scherer G, Lammers M, Hänsch R, Zachgo S, Buschmann H (2018) Transient genetic transformation of *Mougeotia scalaris* (Zygnematophyceae) mediated by the endogenous α -tubulin1 promoter. *J Phycol* 54:840–849
- Renault H, Werck-Reichhart D, Weng J-K (2019) Harnessing lignin evolution for biotechnological applications. *Curr Opin Biotechnol* 56:105–111. <https://doi.org/10.1016/j.copbio.2018.10.011>
- Rippin M, Becker B, Holzinger A (2017) Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant Cell Physiol* 58:2067–2084. <https://doi.org/10.1093/pcp/pcx136>
- Rippin M, Pichrtová M, Arc E et al (2019) Metatranscriptomic and metabolite profiling reveals vertical heterogeneity within a *Zygnema* green algal mat from Svalbard (High Arctic). *Environ Microbiol* 21:4283–4299. <https://doi.org/10.1111/1462-2920.14788>
- Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11:R25. <https://doi.org/10.1186/gb-2010-11-3-r25>
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Schlösser UG (1994) SAG - Sammlung von Algenkulturen at the University of Göttingen Catalogue of Strains 1994. *Botanica Acta* 107(3):113–186. <https://doi.org/10.1111/j.1438-8677.1994.tb00784.x>
- Sierro N, Batty JND, Ouadi S et al (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun* 5:3883. <https://doi.org/10.1038/ncomms4833>
- Suetsuga N, Mittmann F, Wagner G, Hughes J, Wada M (2006) A chimeric photoreceptor gene, *NEOCHROME*, has arisen twice during plant evolution. *Proc Natl Acad Sci U S A* 102:13705–13709
- Sun Y, Harpazi B, Wijerathna-Yapa A et al (2019) A ligand-independent origin of abscisic acid perception. *Proc Natl Acad Sci U S A* 116:24892–24899. <https://doi.org/10.1073/pnas.1914480116>
- von Schwartzberg K, Bornfleth S, Lindner A-C, Hanelt D (2013) The Microalgae and Zygnematophyceae Collection Hamburg (MZCH) — living cultures for research on rare streptophytic algae. *Algol Stud* 142:77–107
- Wagner G, Klein K (1981) Mechanism of chloroplast movement in *Mougeotia*. *Protoplasma* 109:169–185. <https://doi.org/10.1007/BF01287638>
- Wang L, Li Z, Qian W et al (2011) The arabidopsis purple acid phosphatase AtPAP10 is predominantly associated with the root surface and plays an important role in plant tolerance to phosphate limitation. *Plant Physiol* 157:1283–1299. <https://doi.org/10.1104/pp.111.183723>
- Wang S, Li L, Li H et al (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants* 6:95–106. <https://doi.org/10.1038/s41477-019-0560-3>
- Wickett NJ, Mirarab S, Nguyen N et al (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci U S A* 111:E4859–E4868. <https://doi.org/10.1073/pnas.1323926111>
- Wodniok S, Brinkmann H, Glöckner G et al (2011) Origin of land plants: do conjugating green algae hold the key? *BMC Evol Biol* 11:104. <https://doi.org/10.1186/1471-2148-11-104>
- Zimmer A, Lang D, Richardt S et al (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol Gen Genomics* 278:393–402. <https://doi.org/10.1007/s00438-007-0257-6>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.3 Publication III: Environmental gradients reveal stress hubs predating plant terrestrialization

This research paper was published online in the Journal “Nature Plants” in August 2023. The full article as well as supplementary figures and supplementary datasets can be found online:

<https://doi.org/10.1038/s41477-023-01491-0>

Contribution of Janine Fürst-Jansen, first author

J. M. R. Fürst-Jansen planned and performed all pre-experiments to determine optimal experimental conditions and analyzed the morphological as well as photo-physiological results of these experiments. Based on this, she fine-tuned the experiment settings. She planned and performed the main experiments, assisted by T.D., including all culturing steps and absorption, chlorophyll *a* fluorescence measurements and microscopy work, assisted by M.L.. J. M. R. Fürst-Jansen harvested all samples for Illumina sequencing, performed all RNA isolations and quality checked the samples. Additionally, together with D.K., she isolated lipid droplet enriched fractions from *Mesotaenium* and performed proteomics analysis except for LC-MS/MS measurements.

J. M. R. Fürst-Jansen processed the raw data of all absorption and chlorophyll *a* fluorescence measurements and designed, evaluated, and performed in-depth statistical analyses. Based on this analysis she contributed all data in figure 1, most of the statistical analyses therein, and designed parts of the figure. Additionally, she contributed data to parts of figure 6 and helped with the design and finalization of figures 2, 3, 4, and 6 as well as some of the extended data figures; the latter included all figures on physiology and algal phenotypes. She wrote parts of the manuscript and critically revised the manuscript after submission to the journal.

Other contributions

J.d.V. and M.L. conceived the project. J.d.V. coordinated the project with M.M. M.L. provided the plant material. J.M.R.F.-J., T.D., S.S., M.L., and T.P.R. performed the experimental work. A.D. carried out the computational analysis. O.V., J.M.R.F.-J., P.S., T.I., D.K. and G.H.B. performed the proteomics. D.K., C.H. and I.F. performed the lipid profiling. H.B. investigated the cell division patterns. M.H. and U.H. investigated the photomorphogenesis patterns. A.D. and R.S. built the web resources. J.d.V., A.D. and J.M.R.F.-J. contributed to writing the manuscript. J.d.V. organized the manuscript. All authors commented, discussed and provided input on the final manuscript.

Environmental gradients reveal stress hubs pre-dating plant terrestrialization

Received: 25 October 2022

Accepted: 11 July 2023

Published online: 28 August 2023

 Check for updates

Armin Dadras ^{1,14}, Janine M. R. Fürst-Jansen ^{1,2,14}, Tatyana Darienko¹, Denis Krone¹, Patricia Scholz ³, Siqi Sun ⁴, Cornelia Herrfurth ^{3,5}, Tim P. Rieseberg ¹, Iker Irisarri^{1,2,6}, Rasmus Steinkamp¹, Maike Hansen ⁷, Henrik Buschmann⁸, Oliver Valerius⁹, Gerhard H. Braus ⁹, Ute Hoecker ⁷, Ivo Feussner ^{3,5,10}, Marek Mutwil ¹¹, Till Ischebeck ⁴, Sophie de Vries ¹, Maike Lorenz ¹² & Jan de Vries ^{1,2,13} 

Plant terrestrialization brought forth the land plants (embryophytes). Embryophytes account for most of the biomass on land and evolved from streptophyte algae in a singular event. Recent advances have unravelled the first full genomes of the closest algal relatives of land plants; among the first such species was *Mesotaenium endlicherianum*. Here we used fine-combed RNA sequencing in tandem with a photophysiological assessment on *Mesotaenium* exposed to a continuous range of temperature and light cues. Our data establish a grid of 42 different conditions, resulting in 128 transcriptomes and ~1.5 Tbp (~9.9 billion reads) of data to study the combinatory effects of stress response using clustering along gradients. *Mesotaenium* shares with land plants major hubs in genetic networks underpinning stress response and acclimation. Our data suggest that lipid droplet formation and plastid and cell wall-derived signals have denominated molecular programmes since more than 600 million years of streptophyte evolution—before plants made their first steps on land.

Plant terrestrialization changed the face of our planet. It gave rise to land plants (Embryophyta), the major constituents of Earth's biomass¹ and founders of the current levels of atmospheric oxygen². Land plants belong to the Streptophyta, a monophyletic group that

includes the paraphyletic freshwater and terrestrial streptophyte algae and the monophyletic land plants. Meticulous phylogenomic efforts have established the relationships of land plants to their algal relatives^{3–6}. These data brought a surprise: the filamentous and

¹Institute of Microbiology and Genetics, Department of Applied Bioinformatics, University of Goettingen, Goettingen, Germany. ²Campus Institute Data Science, University of Goettingen, Goettingen, Germany. ³Albrecht-von-Haller-Institute for Plant Sciences, Department of Plant Biochemistry, University of Goettingen, Goettingen, Germany. ⁴Institute of Plant Biology and Biotechnology, Green Biotechnology, University of Münster, Münster, Germany. ⁵Goettingen Center for Molecular Biosciences, Service Unit for Metabolomics and Lipidomics, University of Goettingen, Goettingen, Germany. ⁶Section Phylogenomics, Centre for Molecular Biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change, Museum of Nature, Hamburg, Germany. ⁷Institute for Plant Sciences and Cluster of Excellence on Plant Sciences, Biocenter, University of Cologne, Cologne, Germany. ⁸Faculty of Applied Computer Sciences and Biosciences, Section Biotechnology and Chemistry, Molecular Biotechnology, University of Applied Sciences Mittweida, Mittweida, Germany. ⁹Institute of Microbiology and Genetics and Göttingen Center for Molecular Biosciences and Service Unit LCMS Protein Analytics, Department of Molecular Microbiology and Genetics, University of Goettingen, Goettingen, Germany. ¹⁰Goettingen Center for Molecular Biosciences, Department of Plant Biochemistry, University of Goettingen, Goettingen, Germany. ¹¹School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. ¹²Albrecht-von-Haller-Institute for Plant Sciences, Department of Experimental Phycology and SAG Culture Collection of Algae, University of Goettingen, Goettingen, Germany. ¹³Goettingen Center for Molecular Biosciences, Department of Applied Bioinformatics, University of Goettingen, Goettingen, Germany. ¹⁴These authors contributed equally: Armin Dadras, Janine M. R. Fürst-Jansen. ✉e-mail: devries.jan@uni-goettingen.de

unicellular Zygnematophyceae—and not other morphologically more elaborate algae—are the closest algal relatives of land plants. Now, the first genomes of major orders of Zygnematophyceae (see ref. 7) are at hand: *Mesotaenium endlicherianum*⁸, *Spirogloea muscicola*⁸, *Zygnema circumcarinatum*⁹, *Closterium peracerosum–strigosum–littorale*¹⁰ and *Penium margaritaceum*¹¹. Using these, we are beginning to redefine the molecular chassis shared by land plants and their closest algal relatives. Included in this shared chassis will be those genes that facilitated plant terrestrialization. In this Article, we focus on one critical aspect: the molecular toolkit for the response to environmental challenges. For this, we used the unicellular freshwater/subaerial alga *Mesotaenium endlicherianum*.

Land plants use a multi-layered system for the adequate response to environmental cues. This involves sensing, signalling and response, mainly by the production of, for example, protective compounds. Some of the most versatile patterns in land plant genome evolution concern genes for environmental adaptation^{12–14}. That said, there is a shared core of key regulatory and response factors that are at the heart of plant physiology. These include phytohormones such as abscisic acid (ABA) found in non-vascular and vascular plants^{15,16}, protective compounds resting on specialized metabolic routes such as phenylpropanoid-derived compounds and proteins such as LATE EMBRYOGENESIS ABUNDANT (LEA)^{17,18}. Many of the genes integrated into these stress-relevant metabolic routes have homologues in streptophyte algae¹⁹. Taking angiosperms as reference, such stress-relevant pathways are often patchy. Whether these are also used under the relevant conditions is currently unknown. For example, while Zygnematophyceae have a homologue to the ABA-receptor PYL^{8,20}, this homologue works in a different, ABA-independent fashion²¹. Thus, it is important to put the genetic chassis that could act under environmental shifts to the test.

Here we used a fine grid of a bifactorial gradient for two key terrestrial stressors, variation in irradiance and temperature, to probe the genetic network that the closest algal relatives of land plants possess for the responsiveness to abiotic cues. Correlating environmental parameters, physiology and global differential gene expression patterns from 128 transcriptomes (9,892,511,114 reads, 1.5 Tbp of data) across 126 distinct samples covering a temperature range of >20 °C and light range of >500 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, we pinpoint hubs in the circuits that have been shared along more than 600 million years of streptophyte evolution.

Results

A physiological grid: co-dependency of eurythermy and euryphyty

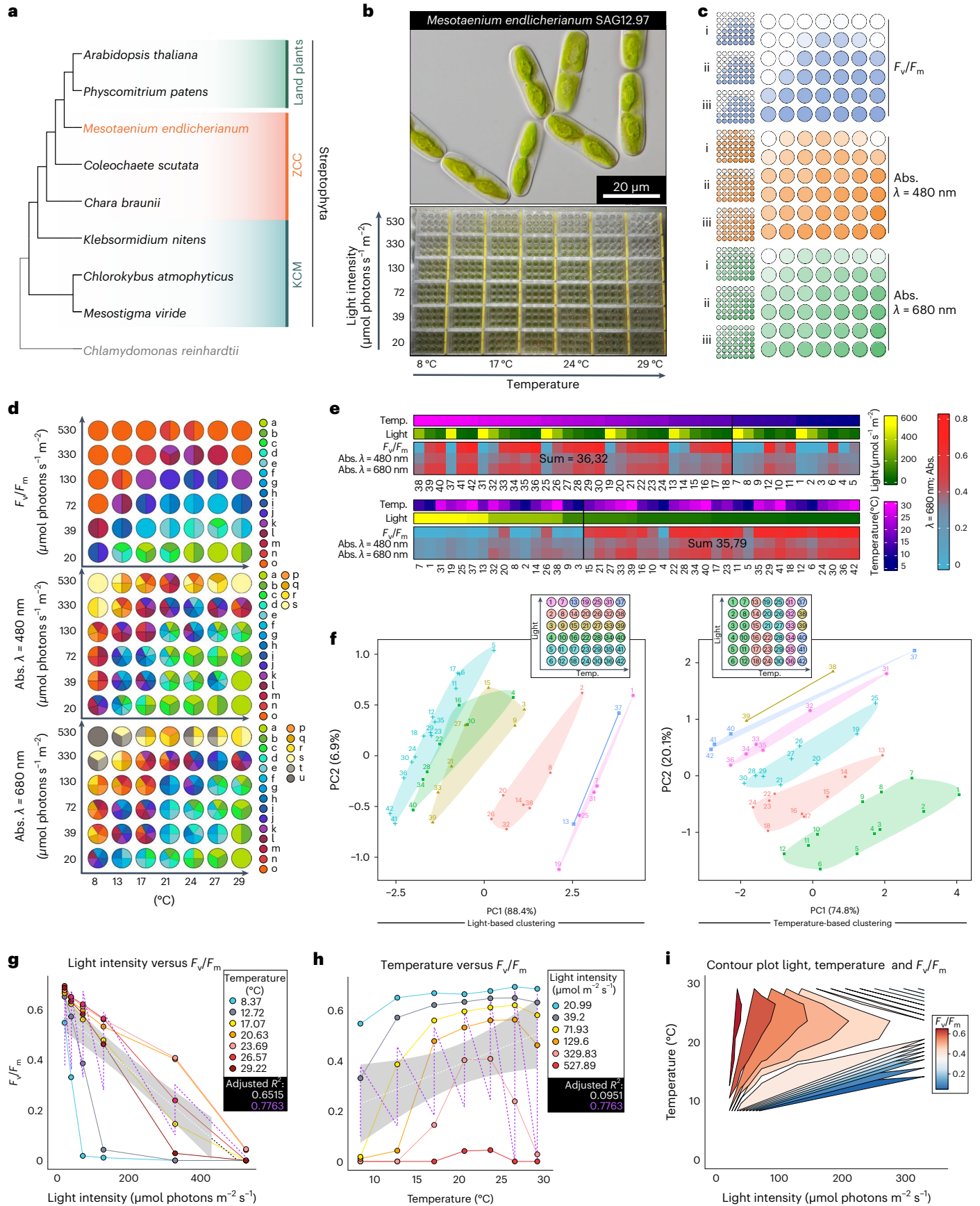
We studied the genome-sequenced strain SAG 12.97 of the freshwater alga *Mesotaenium endlicherianum*, a member of the Zygnematophyceae, the closest algal relatives of land plants⁸ (Fig. 1a,b). Natural habitats for *Mesotaenium*, belonging to the order Serritaeniales, are diverse—ranging from plankton to aeroterrestrial^{7,8}. We cultivated

Mesotaenium in a large-scale setup in 1.5 l of C-medium up to a cell density of 0.33 AU at 680 nm and distributed the culture across 504 wells (42 12-well plates, 2.5 ml of culture per well). Well plates were placed on a table with a temperature gradient from 8.6 ± 0.5 °C to 29.2 ± 0.5 °C on the x axis; from above, light-emitting diode (LED) lamps created an irradiance gradient from 21.0 ± 2.0 to 527.9 ± 14.0 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ across the y axis, thus creating a two-dimensional gradient table (Fig. 1b, Supplementary Table 1 and for light quality, see Extended Data Fig. 4); the conditions were chosen to strike a balance between cell viability and environmental challenge, as determined in a set of pre-experiments (Extended Data Figs. 4–6 and Methods). The 504 cultures were exposed to this gradient setup for 65 h. The physiological status of the algae was assessed by determining the maximum quantum yield of photosystem II (F_v/F_m) using pulse amplitude modulation fluorometry (IMAGING-PAM, Walz) and a microplate reader with absorption at 480, 680 and 750 nm (Fig. 1c, Extended Data Fig. 5 and Supplementary Fig. 1a); the entire procedure was repeated in three successive biological replicates (that is, three runs of the table, 504 F_v/F_m and 4,536 absorption measurements per replicate).

The algae showed significant differences ($P \leq 0.001$) in F_v/F_m values as well as absorption values, both decreasing with rising intensities of irradiance (for F_v/F_m values at 20.5 ± 1.0 °C: from 0.66 ± 0.02 at a light intensity of 21.14 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ to 0.042 ± 0.04 at a light intensity of 534.7 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) (Fig. 1d, Supplementary Fig. 1 and Supplementary Table 2); despite ample growth at 29.2 ± 0.5 °C and low irradiance, higher temperatures (that is, above 29 °C) were out of the tolerable scope of *Mesotaenium* (Extended Data Fig. 5). We recorded the lowest F_v/F_m values (down to zero) at conditions of highest irradiance and lowest temperature. Under the ranges tested here, the low temperature had a stronger negative impact on physiology than light. For example, F_v/F_m values at 8.6 ± 0.5 °C and 133 ± 27 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ are in a different significance group ($P \leq 0.001$) (group o in Fig. 1d) than F_v/F_m values at 29.2 ± 0.5 °C at 118 ± 25 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ (purple, group k in Fig. 1d). Values on physiology clustered by light were less broadly distributed than if clustered by temperature (Fig. 1e,f). Even the highest light intensity (527.9 ± 14.0 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) was stressful but tolerable for the physiology of *Mesotaenium* at temperatures between 20.5 ± 0.1 °C ($F_v/F_m = 0.042 \pm 0.04$) and 25.3 ± 0.1 °C ($F_v/F_m = 0.045 \pm 0.04$); more extreme temperatures resulted in undetectable F_v/F_m values. On the basis of the environmental parameters tested herein, eurythermy (broad viable tolerance of temperature) might establish the foundation for euryphyty (broad viable tolerance of light intensities) in *M. endlicherianum*. Thus, we used regression analysis to understand the effect and importance of the independent values of light and temperature on the dependent physiological values (Fig. 1g–i, Supplementary Fig. 1b and Supplementary Table 2). We find that physiology was always better explained by a combination of temperature and light than a single parameter alone (for example, for F_v/F_m , R^2 of 0.776 versus 0.652 and 0.095; Fig. 1g–i).

Fig. 1 | A fine-combed setup for assessing environmental responses in *Mesotaenium*. **a**, Cladogram of Streptophyta, highlighting that *Mesotaenium endlicherianum* SAG 12.97 is a representative of the closest algal relatives of land plants. KCM, the grade of Klebsormidiophyceae, Chlorokybophyceae and Mesostigmatophyceae; ZCC, the grade of Zygnematophyceae, Coleochaetophyceae and Charophyceae. **b**, *M. endlicherianum* grown in C-medium in 42 12-well plates on a gradient table that produces a temperature range of 8.6 ± 0.5 °C to 29.2 ± 0.5 °C on the x axis and an irradiance gradient of 21.0 ± 2.0 to 527.9 ± 14.0 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ on the y axis; for phenotyping per well, at least ten micrographs were taken, all showing similar phenotypes of the cells. **c**, Overview of the measured maximum quantum yield F_v/F_m as a proxy for gross physiology (blue) and absorption (abs.) at 480 (orange) and 680 nm (green); individual replicates of the biological triplicates are shown on the left and the average values are shown on the right. **d**, Statistical analysis of the physiological values (F_v/F_m , abs. 480 nm, abs. 680 nm). Numbers correspond

to environmental conditions on the table. Biological triplicates were grouped into significant groups (a–o, a–s and a–u) with R (version 4.1.3) using a Kruskal–Wallis test coupled with Fisher’s least significance; P values were Bonferroni corrected. Significant differences at $P \leq 0.001$ are shown as letters. **e**, Heat maps displaying averaged physiological values of the 42 conditions sorted either by temperature (temp.) or light. A cut-off was set (black vertical line) on the basis of the distribution of the highest values, which were then summed to determine a positive correlation with temperature or light conditions. **f**, Two PCAs showing the correlation of light conditions (left) or temperature conditions (right) to physiological values (F_v/F_m , abs. 480, 680 nm). Clusters are shown in different colours, which are also visualized in an overview scheme of the gradient table at the top of the plots. **g,h**, Unifactorial regression analysis of light intensity (**g**) and temperature (**h**) versus F_v/F_m ; note the unifactorial linear regression curves (white) versus the bifactorial (violet). **i**, Contour plot of the bifactorial impact of light and temperature on F_v/F_m (gradient colour).



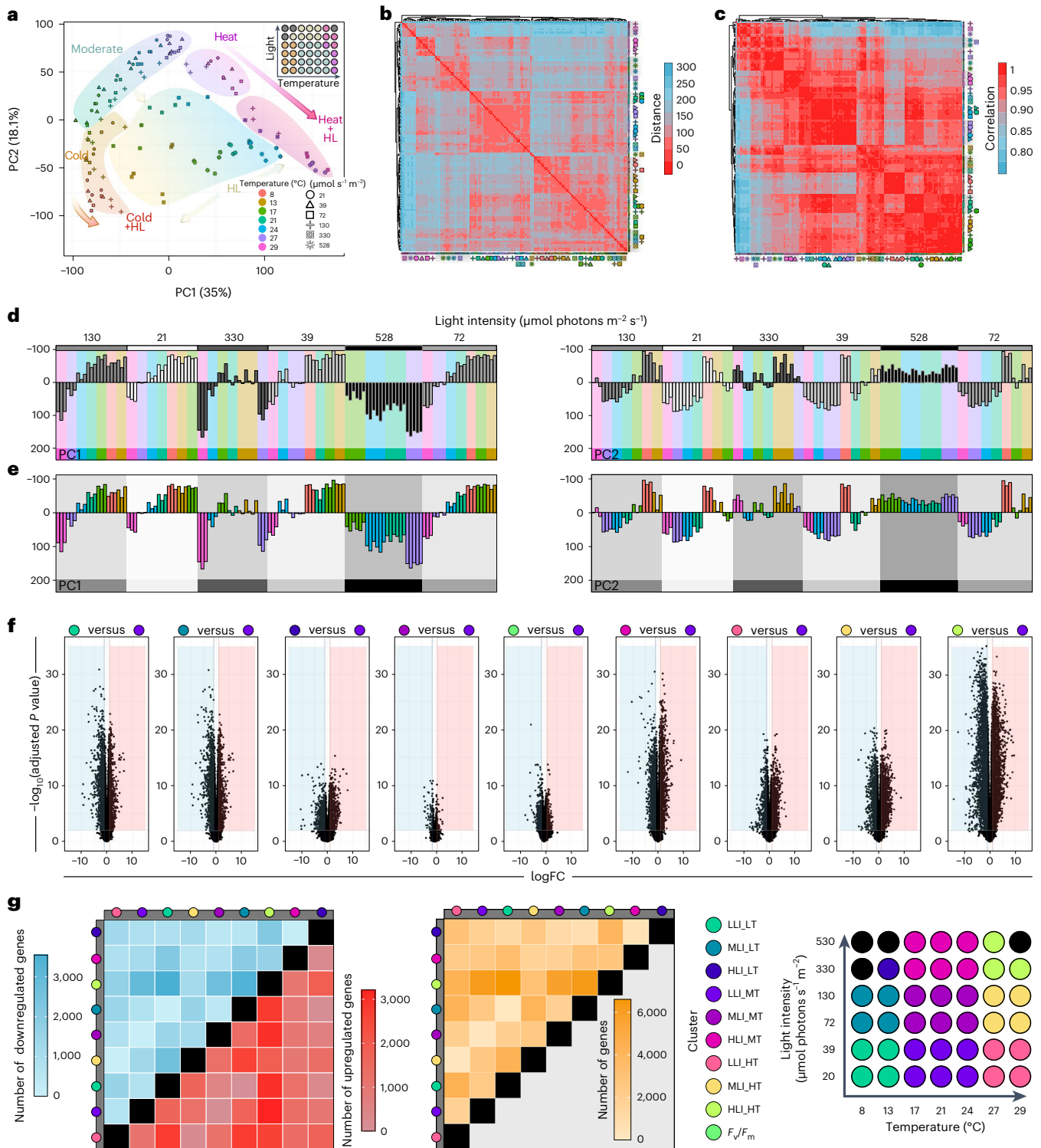


Fig. 2 | Global profiles of environment-governed gene expression response.

a, PCA visualizing PC1 and PC2. Backgrounds were drawn to highlight our interpretation of the observed trends; samples are coded by colour (temperature) and symbols (irradiance in $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Samples that did not yield usable RNA are indicated as grey dots in the top-right overview of the experimental setup. **b**, Visualization of Euclidean distances between samples via heat map, from red, zero distance, to blue, furthest distance (a distance of 300). **c**, Heat map of Spearman correlation between samples, from red, maximum correlation (1.0), to blue, least correlation (<0.8). The clusters were calculated via the Euclidean distance. **d, e**, PC1 and PC2 scrutinized using a small multiples method of light intensity (**d**) and temperature (**e**). In **d**, shades of grey correspond to different light intensities. In **e**, different

colours represent different temperatures and were mapped with the same colours as **a**. To perform differential gene expression analysis, we divided the table into nine sectors (see scheme of the table); additionally, a tenth group was raised based on $F_v/F_m < 0.5$. Linear models were fitted for each gene and empirical Bayes statistics computed for DEGs by the limma package. In total, 37 comparisons were made. DEGs were defined as genes with an absolute fold change (FC) ≥ 2 and Benjamini–Hochberg-adjusted P value less than 0.01. **f**, Volcano plots of DEGs for nine selected comparisons based on the sectors and the $F_v/F_m < 0.5$ criterion. **g**, Heat maps of numbers of DEGs for all sector-based comparisons (blue, downregulation; red, upregulation; yellow, sum of up- and downregulated genes); grey bars label the first component (treatment) for calculating the contrasts (treatment versus control).

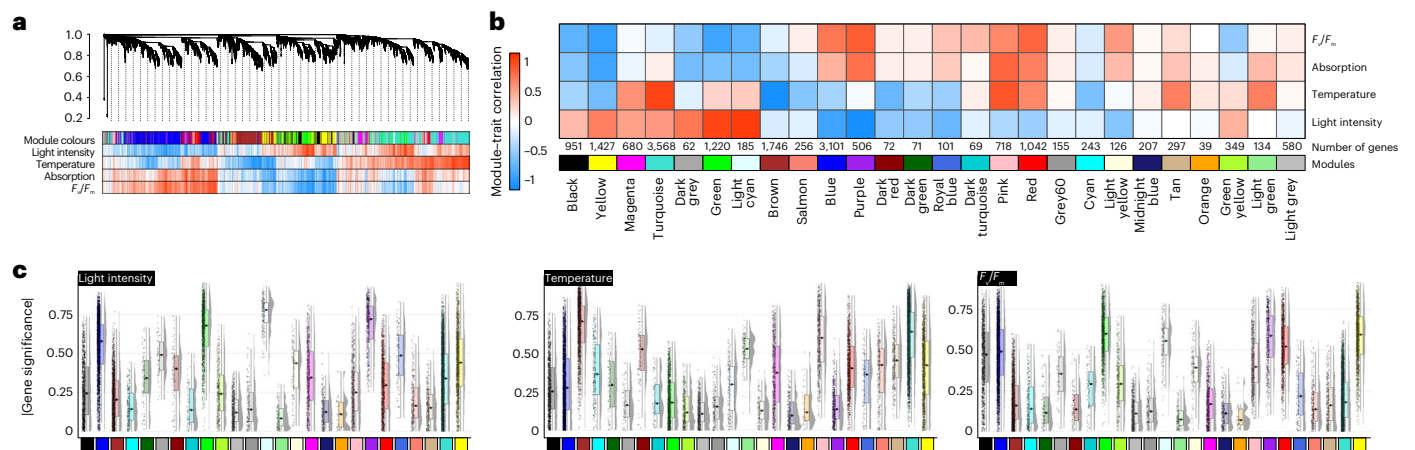


Fig. 4 | Unsupervised gene expression clusters recover genetic programmes separated by environmental cues. Gene expression clustering into 26 coloured modules was performed using WGCNA; grey is the module of unclustered genes. **a**, Hierarchical cluster tree of 17,095 genes. The heat map below the dendrogram and module colour assignment shows the gene significance measure (from red, positive correlation, to white, no correlation, to blue, negative correlation) for the four different conditions or physiological parameters. **b**, Heat map of the module–trait correlation based on eigengenes (from red, positive correlation, to white, no correlation, to blue, negative correlation); see Supplementary

Fig. 7. **c**, Box plots of the mean gene significance across modules (given in the corresponding module colour) towards the parameters light intensity, temperature and F_v/F_m . The box plots display the interquartile range (IQR) of the data, compactly displaying the distribution of a continuous variable. They visualize five summary statistics (the median, two hinges and two whiskers). The upper whiskers extends from the hinges to the largest/smallest value no further than $1.5 \times$ IQR from the hinge. Each data point (n) is a gene, and the total n of genes is the same as shown in **b**. We calculated the gene significance for each gene using the WGCNA package and Pearson method.

set of evidence used to calculate AED and it shows higher congruence with them (Supplementary Fig. 3a). Thus, we pseudo-aligned our data onto the new *Mesotaenium* transcriptome V2 (average alignment rate was 87.31%; Supplementary Table 4a).

Cheng et al.³ reported that 33.2% of the genome was impacted by transposable elements. We surveyed V2 for protein domains related to transposon biology, retrieving 6,186 entries in 1,748 unique genes (Supplementary Table 4b). Among the 96 that passed the expression threshold, high temperature (29 °C) appeared to have had the strongest effect on transposable element mobilization (Supplementary Fig. 3b).

To understand the gross profile of the gene expression data, we performed a principal component analysis (PCA; Fig. 2a). Independent biological replicates from the same condition clustered in close proximity. High temperature followed by irradiance brought forth clear separation of the data, with PC1 describing 35% and PC2 describing 18.1% of the variance. We evaluated the distance (Fig. 2b) and Spearman correlation (Fig. 2c) using all genes to look for trends among different environmental conditions. The data can be grouped into at least three categories: (1) samples with high light and/or high temperature, (2) a collection of low-temperature (8, 13 and 17 °C) samples, and (3) samples at moderate conditions. Large clusters included low to medium light + medium temperature (‘moderate’ conditions), high light + high temperature, and high light (Fig. 2a). Most distinct was the cluster formed by samples from the high temperature + high light (small multiples; Fig. 2d,e).

Plastid-related genes stand out in differential gene expression profiles

For dissecting the differential gene expression responses, we divided the table into nine sectors and, additionally, a cohort of stressed algae based on $F_v/F_m < 0.5$ (Fig. 2f,g). We performed 36 comparisons, among which we focused on nine, which additionally included the F_v/F_m -based comparison. Genes were considered to be differentially expressed between groups at an absolute fold change ≥ 2 and a Benjamini–Hochberg-corrected $P \leq 0.01$ (Fig. 2f,g). The intensity of environmental cues governed gross gene expression profiles as increasing disparity between conditions yielded more differentially expressed genes (DEGs), generally following the pattern of the PCA (compare

Fig. 2a,g). The most differentially regulated genes (6,578) were pinpointed by comparing low light and low temperature (LLI_LT) versus high light and high temperature (HLI_HT). Enriched Gene Ontology (GO) terms among regulated genes most frequently included plastid biology-associated genes (Extended Data Fig. 1); similar patterns were recovered in 63 unifactorial comparisons where we kept one environmental parameter constant (Extended Data Fig. 7a). To scrutinize our data for specific genes that show a robust and universal response to alterations in the environment, we intersected all 8,157 significantly regulated genes pinpointed by the different comparisons: 3, 30 and 124 genes overlapped among all 9, 8 and 7 comparisons, respectively. These concertedly pinpointed genes were mostly light harvesting genes, corroborating the importance of plastids in the overall cell biology of *Mesotaenium* (Extended Data Fig. 7b). Indeed, the 30 genes found in all comparisons included, for example, reactive oxygen species (ROS)-relevant genes such as *EARLY LIGHT-INDUCIBLE PROTEIN (ELIP)* and fatty acid metabolic genes.

How do these responses compare across land plants’ close relatives? To answer this, we downloaded major stress transcriptome data from streptophyte algae^{9,20,22–25}, inferred significant differential gene expression between stress treatment and control per species, and asked whether regulated genes belong to the same phylogenetic hierarchical orthogroups (HOGs, inferred with Orthofinder²⁶). Depending on the phylogenetic distance of the species, we found between 3,107 and 6,458 HOGs shared with *Mesotaenium*, with 46.6–73.0% shared within and 15.8–30.4% outside of the clade Zygnematophyceae (Fig. 3a). Of these shared HOGs, between 4.6% and 59.8% show shared regulation with *Mesotaenium*. The degree of similarity depends on treatment not phylogenetic position. The most common responses across species were related to chloroplasts and photosynthesis (Fig. 3b and Extended Data Fig. 2). However, within Zygnematophyceae, additional signalling processes such as kinase activities and calcium-dependent signalling stood out (Fig. 3b and Extended Data Fig. 2), corroborating (1) their noted importance in Zygnematophyceae^{23,24} and (2) the concept that important steps in the evolution of streptophyte calcium signalling system (Extended Data Fig. 2) occurred before plant terrestrialization²⁷.

To understand whether these genes integrate into the context of molecular programmes, we next analysed gene co-expression.

Gene expression clusters recover ancient genetic programmes

The environmental gradients triggered changes in the expression of gene cohorts. We wanted to understand their concerted action independent of any prioritization guided by homology to any land plant genes—solely from the molecular programmes that operated in the algae. To do so, we applied a weighted gene co-expression network analysis²⁸ (WGCNA) for unsupervised clustering (Fig. 4, Supplementary Figs. 4–7 and 10–13 and Extended Data Fig. 8). To then understand the driving forces behind these changes, we turned to the highly connected genes (nodes) in the network (hubs) (Fig. 5).

We clustered the 17,905 genes expressed in our samples (passing the minimum expression threshold) into 26 modules, which we refer to with colours (Fig. 4a). Orange is the smallest module (39 genes), and the largest modules are turquoise, blue and brown with 3,568, 3,101 and 1,746 genes, respectively (Fig. 4b). The samples reflect a range of distinct physiological conditions and resulting data are a combined expression of the different environmental cues and the modulation of the algal physiology. To investigate the biological role of each module, we used their eigengenes as representatives for the modules' gene expression profiles and correlated their behaviour with the two environmental cues (light intensity and temperature), as well as the physiological parameters absorption and F_v/F_m (Fig. 4b,c). One of the foremost general patterns in cellular response to stress are ROS. ROS act as signals as well as culprits that, if not quenched, damage biomolecules; GO terms capture ROS biology (Extended Data Fig. 3), especially in module green that positively correlates with light intensity ($r = 0.88, P = 6 \times 10^{-43}$) and negatively with F_v/F_m ($r = -0.79, P = 6 \times 10^{-29}$) (Extended Data Fig. 3, Supplementary Figs. 4–7 and Supplementary Tables 5 and 6).

The clusters also recover the genetic signatures of thriving algae. Module purple negatively correlates with increasing light ($r = -0.94, P = 3 \times 10^{-60}$) and positively with absorption and F_v/F_m ($r = 0.71, P = 3 \times 10^{-21}$ and $r = 0.79, P = 4 \times 10^{-28}$). These dense and physiologically healthy cell populations (experiencing no light stress) likely ramped up cell division (Extended Data Fig. 3 and Supplementary Table 6), signified by homologues of cyclin and *TPX2* appearing as hub genes (Fig. 5e). The ninth most connected hub is a kinesin homologous to genes coding for proteins such as PHRAGMOPLAST ORIENTING KINESIN 2, and homologues of the important growth regulators²⁹ Tesmin and TSO1 ranked at positions 7, 15 and 17 of the most connected hubs in module purple (Fig. 5e and Supplementary Table 7).

To understand the evolutionary conservation of the genetic programmes in these modules, we processed 212 publicly available RNA-seq datasets from *Zygnema circumcarinatum*⁹, *M. polymorpha*, *P. patens* and *A. thaliana* exposed to diverse abiotic challenges using the same WGCNA pipeline, which yielded between 12 and 29 modules. We determined orthogroups between the modules of these different species and compared the similarity in modules by calculating Jaccard

indices (Fig. 5f) and GO-term enrichment in these modules (Fig. 5g and Supplementary Fig. 14b–e). Also here, blue, brown, turquoise and yellow stand out as important and likely conserved environmental response modules (compare Figs. 4b and 5f and Extended Data Fig. 3). We further analysed shared connectivity of hub orthogroups. For the mentioned Tesmin and *TSO1* orthogroups (Fig. 5h), this reveals that they are likely connected regulators of cell division since about 600 million years of streptophyte evolution. To scrutinize this aspect, we inferred the evolutionary history of the 160 hubs using maximum likelihood phylogenetic analyses (Fig. 5h; data on Zenodo). We retrieved 135 phylogenies, 107 of which indicate that the hubs are in gene families that were present in (or before) the last common ancestor of Zygnematophyceae and land plants. Thus, they pre-date plant terrestrialization.

Conserved hubs: integration of plastid and cell physiology

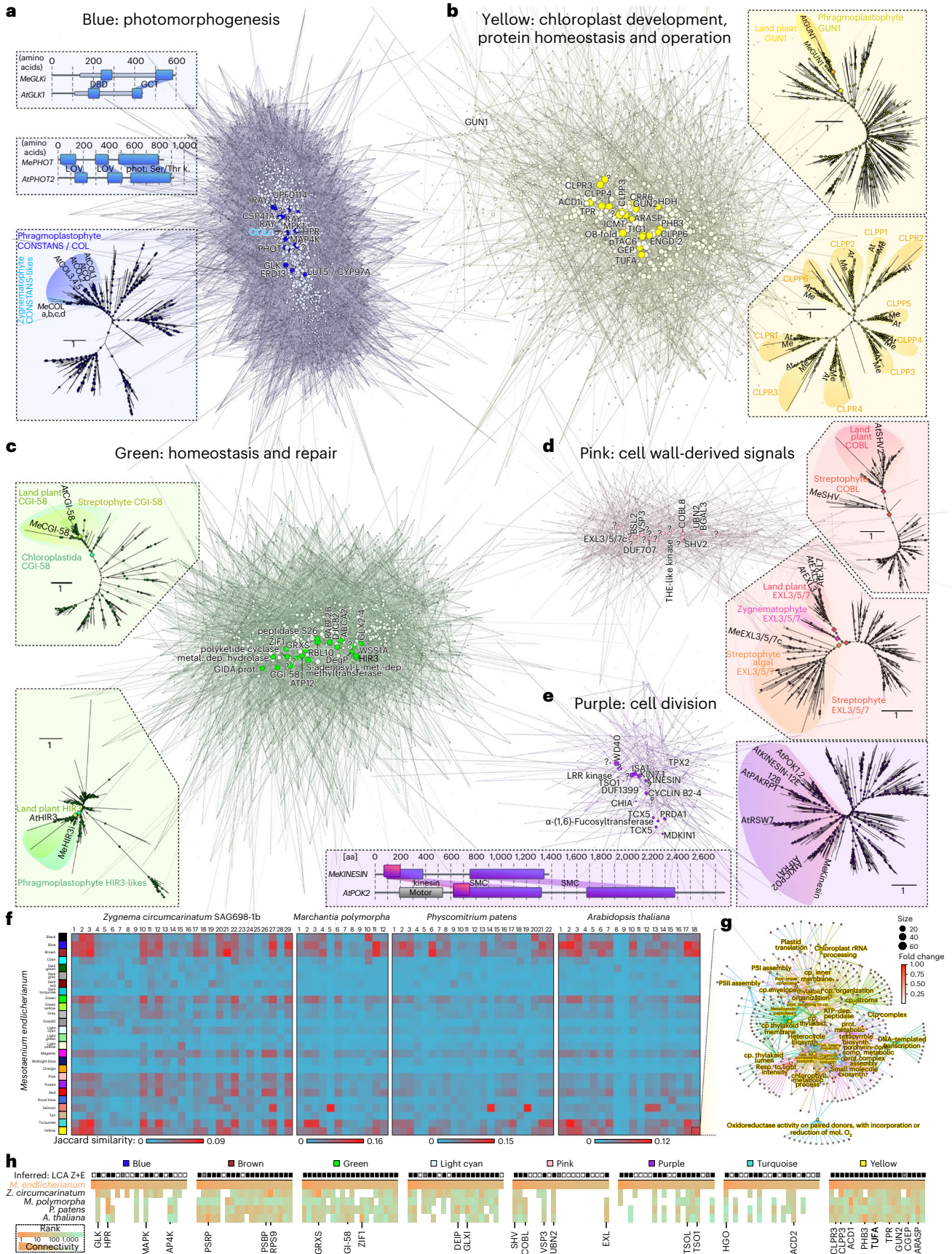
Chloroplasts act as environmental sensors in land plant cells³⁰. In concert with this, many of the modules we identified were associated with plastid biology and/or physiology (Extended Data Fig. 3, Supplementary Figs. 4–7 and Supplementary Table 6). Module brown is enriched in GO terms related to plastids, general transcription and translation, and negatively correlates with temperature ($r = -0.95, P = 7 \times 10^{-65}$; Extended Data Fig. 3 and Supplementary Fig. 5). Among the top 20 hub genes in module brown, 12 are associated with translation and ribosomes (Supplementary Table 7). As expected, this cluster shows conservation in enriched functions of its related modules in the other four streptophytes, including shared high connectivity of hubs (Fig. 5f,h). The module light cyan positively correlates with increasing light ($r = 0.93, P = 1 \times 10^{-56}$; Supplementary Fig. 6) and negatively with F_v/F_m ($r = -0.67, P = 5 \times 10^{-18}$; Supplementary Fig. 4). It features not only hubs related to ROS homeostasis from the thioredoxin superfamily and other light-induced proteins, but also pigment and apocarotenoid metabolism (Extended Data Fig. 3); these are the source of important signals from the chloroplast that likely have deep evolutionary roots¹⁹ and are also formed by light-dependent oxidative reactions³¹. The blue module negatively correlates with increasing light ($r = -0.76, P = 10^{-25}$) and positively with F_v/F_m ($r = 0.67, P = 2 \times 10^{-18}$). Concomitantly, the blue module has a high number of enriched GO terms, many of which are plastid-related terms, cellular signalling and terms that tie the two together—that is, signalling processes emanating from the plastid (Extended Data Fig. 3 and Supplementary Table 6). Such responses align with similar clusters in other species (Fig. 5f), where the related *Arabidopsis* modules 2 and 10 show terms for light intensity and quality (Supplementary Fig. 14b).

The hubs of many modules, including those in blue, light cyan and yellow mentioned before, reflect an association with plastid-related processes. To highlight a few, the second most connected gene in module blue is a homologue of *GOLDEN2-LIKE1 (GLK1)* (Supplementary Fig. 8). *GLK1* is a transcription factor (TF) that regulates chloroplast

Fig. 5 | Molecular programmes for environmental responses around recurrent plant hubs. a–e.

Visualization of the co-expression network clustered by WGCNA into the modules blue (3,101) (a), yellow (1,427) (b), green (1,220) (c), pink (718) (d) and purple (506 genes) (e). Nodes (circles) represent genes connected by edges whose weight (light to dark colour) is based on a weighted TOM. Brightly coloured nodes represent the 20 most connected genes (hubs) and are annotated based on homology; all other nodes are depicted in the corresponding paler colour. Around the clusters, different protein-coding hub genes are highlighted, giving information such as predicted domain structures or phylogenetic relationships; for fully labelled phylogenies, see Supplementary Fig. 26b. Circles in phylogenies give a scale of the ultrafast bootstrap support values; diamonds indicate high (>90%) support for branches separating highlighted clades. An alignment of *GLK* homologues can be found in Supplementary Fig. 8. f. Using WGCNA, co-expression networks were computed from 212 publicly available RNA-seq datasets from *Z. circumcarinatum*, *M. polymorpha*, *P. patens* and *A. thaliana* exposed to diverse

abiotic challenges, yielding between 12 and 29 modules (labelled above the heat map), and orthogroups for all genes in the modules of these different species were determined. The heat map shows the similarity, based on Jaccard indices, between the modules of *Mesotaenium* (same colours as throughout the paper, see Fig. 4b) and the co-expression modules in the three land plants as well as *Zygnema*; red to blue colour gradients indicate high to low Jaccard similarity. g. Cnet plot of the enriched GO terms in the module 'Arabidopsis 18', which has high Jaccard similarity to the *M. endlicherianum* module yellow—note the recurrent terms of plastid operation and, especially, the Clp complex. h. Heat map of the connectivity ranks across all five species for homologues of hub genes of *Mesotaenium*, from orange (high) to green (low connectivity). Black boxes (top row) indicate if our phylogenies (see data on Zenodo) suggest that the hub genes fall into families that were present in the last common ancestor of Zygnematophyceae and land plants, and hence emerged before plant terrestrialization; white boxes signify the absence of such indication and grey boxes highlight ambiguous relationships.



development and the activity of nuclear genes involved in photosynthetic light reaction and chlorophyll biosynthesis^{32–34}; indeed, genes in the *GLK* orthogroup are highly connected throughout the modules of land plants, and in the zygnematophyte, *Zygnema* a *GLK* homologue is the eighth most connected gene in its module (Fig. 5h). Blue also features hydroxypyruvate reductase-coding gene, important in photorespiration³⁵, as the fourth most connected hub, which appears in the top-five most connected genes in the bryophytes (Fig. 5h). A CYP450 gene homologous to *LUTEIN DEFICIENT 5* (*LUT5*), is the seventh most connected gene, suggesting the involvement of pigment-related signalling. Module 21 in *P. patens* is dominated by ABA signalling (Supplementary Fig. 14d) and it is similar to *Mesotaenium* modules turquoise and blue (Fig. 5f), enriched in homologues of ABA-activated signalling (Extended Data Fig. 3), featuring a highly connected homologue of *ABA-RESPONSIVE ELEMENT-BINDING FACTOR 2* (*ABF2*). Thus, parts of the ABA signalling module consist of ancient wires whose relevance in environmental response pre-date plant terrestrialization, and ABA dependency^{20,21,36}.

Next to *GLK1*—the most connected TF-coding gene—other highly connected transcriptional regulators appear in module blue. These include homologues of photomorphogenesis-regulating genes such as *CONSTANS-like 3* (*COL3*, the fourth most connected TF-coding gene) and *CONSTITUTIVE PHOTOMORPHOGENIC 1* (*COPI*); *CO/COL* and *GLKs* are both degradation targets of *COPI* (refs. 37–39). Further, the circadian regulator⁴⁰ *BROTHER OF LUX ARRHYTHMO* is the second most connected TF-coding gene in module blue. All of this aligns with the similarity to the *Arabidopsis* module 2 and the *P. patens* module 6, featuring, next to light quality, also photoperiodism (Fig. 5f and Supplementary Fig. 14e,f). Further, homologues of *ETHYLENE-INSENSITIVE3-like 1* (the sixth most connected TF-coding gene) and several *ETHYLENE RESPONSE FACTORS* (*ERFs*) are among the most connected TF-coding genes. Previous investigations of the Zygnematophyceae *Spirogyra pratensis* have shown that *SpEIN3* can rescue *Arabidopsis ein3-1* mutant plants⁴¹; exogenous application of ethylene on *Spirogyra* triggers stress-, plastid- and photosynthesis-associated gene expression responses similar to land plants²², which we recover, as outlined, across co-expression modules (Fig. 5f, Extended Data Fig. 3 and Supplementary Fig. 14a–f) and shared differential patterns (Fig. 3b and Extended Data Fig. 2). This speaks for a conserved regulatory framework that involves the plastid, photosynthesis, ethylene-associated factors, and maybe ethylene itself, in environmental signalling cascades in the common ancestor of land plants and their closest algal relatives.

Module yellow correlates positively with light intensity ($r = 0.62$, $P = 10^{-14}$) and negatively with absorption and F_v/F_m ($r = -0.79$, $P = 10^{-28}$ and $r = -0.81$, $P = 3 \times 10^{-31}$; Fig. 4b); GO terms are associated with plastids and proteolytic enzymes^{42,43} (FtsH and ClpP), recapitulating well-known ties of protein homeostasis and plastid maintenance (Extended Data Fig. 3). Indeed, yellow features five hubs that are homologous to genes coding for CLP proteases, critical for chloroplast protein homeostasis^{44,45}, and hubs homologous to genes that orchestrate the coordination of transcriptional activity between chloroplasts and the nucleus (Fig. 5b); the latter includes homologues of (1) *pTAC6*, which is essential for plastid gene expression and thus chloroplast development in *Arabidopsis*⁴⁶, and (2) a homologue of *GENOMES UNCOUPLED 2*, one of the foremost genes in the classical plastid–nucleus communication pathway⁴⁷. Among the TF-coding genes in module yellow is a homologue of the bZIP light signalling master regulator *ELONGATED HYPOCOTYL 5* (ref. 48) (*HYS*). Module yellow is among those with the most consistency in similar modules across the analysed streptophyte co-expression networks and hubs (Fig. 5f,h), as exemplified by the GO term similarities between yellow and *Arabidopsis* module 18 (compare Fig. 5g and Extended Data Fig. 3) and the consistency of the plastid operational genes as hubs (Fig. 5h). Hence, hallmark genes for plastid operation and its integration into molecular cell physiology probably acted in concert since before the dawn of embryophytes.

Of ancient signalling cascades and cell wall perturbation

Mitogen-activated protein kinases (MAPK) constitute environmental response pathways in all eukaryotes⁴⁹. In land plants, several abiotic and biotic cues have been described to trigger MAPK-mediated signalling^{50–53}. Genes coding for MAPK and phototropin kinases appear as hubs in module blue. Moreover, plant MAPK-based signalling is interwoven with wound response and brassinosteroid signalling⁵⁰; the *MAPK* orthologue in *Zygnema* is also highly connected (Fig. 5h) and blue is similar to the kinase-rich module 17 of *Arabidopsis* (Fig. 5f). Stress often coincides with a perturbation of plant cell wall homeostasis. Module pink includes hubs for such wounding and cell wall-derived signals. This pairs with the GO term brassinosteroid signalling, which balances growth, cell wall homeostasis and stress in *Arabidopsis*^{54,55}. Among the hubs in pink are homologues for (1) diverse receptor kinases known from *Arabidopsis* to sense alterations in cell wall integrity⁵⁶ and (2) *EXORDIUM-like* (*EXL*; *Mesotaenium* has 12 *EXL* homologues), which integrates growth with environmental signalling⁵⁷ (Fig. 5d). This pairs with genes coding for the COBRA family proteins being the most and third most connected hubs in the module. COBRA proteins are known to be involved in cell expansion and balancing pathogen response with growth^{58–60}. It appears that *Mesotaenium* bears parts of a loop that senses physico-chemical perturbation of cell wall homeostasis; in land plants, these loops include brassinosteroid signalling⁶¹ and wiring of the core genes mentioned here are ancient, evident by the recurrent high connectivity of *EXL* and *COBL* homologues (Fig. 5h) throughout 600 million years of streptophyte evolution.

LDs: a response pre-dating plant terrestrialization

In land plants, lipid droplet (LD) formation and triacylglycerol (TAG) accumulation are common to many stress responses, including heat, cold and drought^{62–66}. We observed that cells of *Mesotaenium* accumulated inclusions resembling LDs upon prolonged exposure to stress (Fig. 6a). Consistently, these globular structures were stained by BODIPY 493/503 (EM/EX), a common dye for lipid- and oil-rich compartments^{67,68}. Under different temperature and light conditions, counts of LDs per cell showed significant differences (Fig. 6b and Supplementary Table 8). A *CGI-58* homologue is the tenth most connected hub in module green (Fig. 5c). *CGI-58* is key to lipid homeostasis, causing, if perturbed, Chanarin–Dorfman syndrome in humans and LD overaccumulation in *Arabidopsis*^{69,70} (Fig. 5c); *CGI-58* is the 22nd most connected gene in *Arabidopsis* module 5 (Fig. 5h). Further, differential gene expression profiles pinpointed elevation of transcripts for characteristic LD protein homologues such as steroleosin (HSD1) and oleosin (OLE7) under high temperature and moderate light conditions (29 °C, 21–130 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) and LD-associated protein (LDAP) and PUX10 under high temperature and light conditions (21–29 °C, 130–528 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$; Fig. 6c).

To scrutinize whether these structures are comparable to LDs of land plants, we performed subcellular fractionizations, obtained lipid-rich phases and subjected them to proteomics using liquid chromatography–mass spectrometry (LC–MS). We identified 739 proteins in the putative LD fraction and 1,574 proteins in the total extract (TE) (Supplementary Table 9). Of these, 14 were significantly enriched in the putative LD fraction (Fig. 6d) including hallmark LD proteins⁷¹ such as OLE, caleosin (CLO), HSD and LDAP (Fig. 6e). We confirmed the localization to LDs for these four proteins by transiently expressing mCherry-tagged variants in tobacco pollen tubes; mCherry clearly overlapped with BODIPY 493/503 fluorescence (Fig. 6f). Resembling LDs of seeds⁷¹, we found predominantly TAG in the lipid content of the LDs (Fig. 6g and Supplementary Fig. 27b); the lipid profiles of *Mesotaenium* LDs varied with age of the cultures (Fig. 6h,i). Overall, *Mesotaenium* responds to stress conditions by formation of LDs containing signature proteins typical of embryophytic LDs.

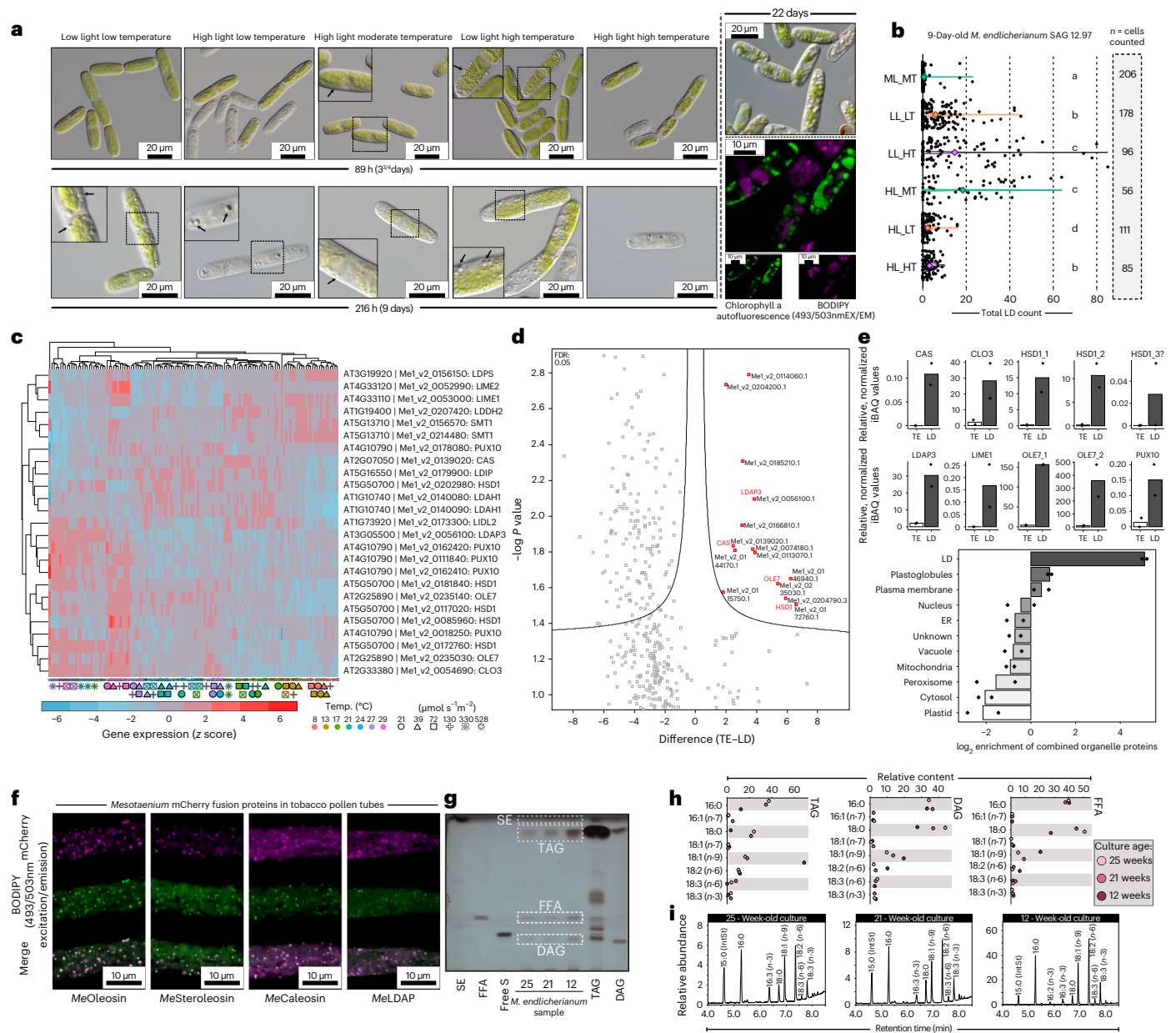


Fig. 6 | LDs accumulate in *Mesotaenium* upon changing environments. **a**, DIC and confocal micrographs of *Mesotaenium endlicherianum* SAG 12.97 cells accumulating LDs (arrows) upon exposure to different temperature/light conditions (abbreviations) of the gradient table for 89 h or 216 h. For confocal microscopy, algae were cultured independent of table conditions at 75 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ and 22 °C for 22 days. LDs are visible as distinct globular structures and were stained with BODIPY (false-coloured green; 493 nm excitation, 503 nm emission); chlorophyll autofluorescence in false-coloured purple; for each condition, at least ten micrographs were taken, all showing similar phenotypes of the cells. **b**, Violin plots of LD quantification after 9 days of exposure to different environmental conditions; significance grouping (Mann–Whitney *U*) is based on $P < 0.05$; see also Supplementary Fig. 27. **c**, Heat map of row-scaled z scores of the expression of homologues for LD biogenesis and function (see also Supplementary Fig. 28). Conditions are displayed at the bottom as symbols in different colours; best *Arabidopsis* hits (via BLASTp) are shown on the right. **d, e**, Proteomic investigation into lipid-enriched phases

extracted from *Mesotaenium*; note the enrichment in hallmark proteins of LDs. Volcano plot showing significantly (false discovery rate (FDR) < 0.05) enriched *Mesotaenium* proteins in the lipid-enriched (LD) versus the TE (**d**). Bar plots show the relative, normalized iBAQ values for ten LD signature proteins detected in *Mesotaenium* (**e**). Bottom bar plot shows the log₂ enrichment of proteins characteristic for subcellular compartments. LL, ML and HL, low, moderate and high light; LT, MT and HT, low, moderate and high temperature, respectively; ER, endoplasmic reticulum. **f**, LD proteins of *M. endlicherianum* localize to LDs in tobacco pollen tubes: cLSM images of transiently expressed proteins appended to mCherry in transiently transformed *N. tabacum* pollen tubes. LDs were stained with BODIPY 493/503; for each construct, the images are representative of at least nine micrographs of transformed pollen tubes per fusion construct. Scale bars, 10 μm . **g, h**, Lipid composition in *M. endlicherianum* LDs of 12- to 25-week-old cultures and standards for sterol esters (SE), FFA, free sterols (free S), TAG and DAG via analytical TLC (**g**) and preparative TLC followed by GC for profiling (**h**). **i**, Full lipid profiles assessed via GC.

Discussion

Owing to their plain morphology, Zygnematophyceae emerged as the unexpected closest algal relatives of land plants^{4–7}. That said, the molecular programmes of Zygnematophyceae speak of their close

relationship to land plants. These point to a conserved chassis that probably operated in the last common ancestor of land plants and algae, featuring the proposed action of various hallmark proteins (for example, PYL homologues²⁰, GRAS family TFs⁸ and more) that were

once considered land plant innovations. During plant terrestrialization, challenges did not come in isolation. The aim of this work was to define stress responses to temperature and irradiance combinations in a close algal relative of land plants. In the approach we have chosen, we made sure to capture both tolerable ranges of cues and those that go beyond tipping points, to allow for robust definition of where stress starts and to pinpointing molecular programmes whose expression dynamics follow the kinetics of their environmental trigger (for example, light intensity in the case of programmes for high light response); this included capturing the well-known double assault of low temperature and high light on the photosynthesis machinery. Building on the genomic resources for *Mesotaenium*, we have here delved into the molecular physiology and genetic programmes of this alga, revealing which programmes bear out when challenged with environmental cues.

Recent studies have proposed homology for the chassis of plastid–nucleus communication upon adverse environmental conditions between land plants and phragmoplastophytic streptophyte algae^{20,72,73}. The GUN pathway probably has a conserved role in chloroplast transcription and streptophyte algal *GUN1* homologues can rescue chloroplast retrograde signalling of *Arabidopsis Atgun1* mutants⁷⁴; the degree of evolutionary conservation in the retrograde signalling pathway across streptophytes remains obscure⁷⁴. Signals from damaged chloroplasts inhibit *GLK1* expression in *Arabidopsis*⁷⁵. The negative correlation of module blue (featuring *MeGLK*) with high light (leading to damaged chloroplasts) supports a role of *MeGLK* in operational retrograde signalling. Indeed, our comparative analyses revealed a consistency in plastidial integration on the basis of similar networks in land plants and *Zygnema* (Fig. 5f) and with regard to highly connected hub genes associated with ROS and plastid protein homeostasis (Fig. 5g,h). Altogether, these insights point to operational processes of the plastid of the closest relatives of land plants, governed by nuclear gene expression for dealing with light regimes and adjustment of photosynthetic performance. On balance, our data underscore that the wires between components in plastid–nucleus communication are probably shared across more than 600 million years of streptophyte evolution.

In land plants, the formation of LDs is known to occur under a variety of adverse environmental conditions^{63,64,76}. Stress-dependent formation of LDs probably evolved before land plants came to be^{24,77,78}, but their molecular underpinnings outside of land plants remain unclear. Here, we confirmed the identity of these *Mesotaenium* LDs using confocal microscopy, LD-specific staining and proteomics. Our comprehensive transcriptomic data illuminate co-expressed modules that might constitute a homologous programme for stress-dependent LDs that acted before plants conquered land.

Methods

Algal culturing and gradient table setup

We used the axenic and genome-sequenced *Mesotaenium endlicherianum* SAG 12.97 (ref. 79) from the Algal Culture Collection, Göttingen (SAG⁸⁰). *Mesotaenium* was cultivated in C-medium⁸¹ for an average of 12 days in an aerated culture glass flask (SCHOTT) at 80 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ (12h/12h light/dark cycle (light from 6 am to 6 pm, Central European winter time) at 18 °C). Before the experiment, cell density was analysed using a LUNA Automated Cell Counter (Logos Biosystems) and set to 2.03×10^7 cells ml^{-1} (diluted with C-medium if needed; settings for cell counting: cell roundness, 60%; minimum size, 3 μm ; maximum size, 60 μm), corresponding to $\text{Abs}_{680\text{nm}} = 0.33$ (Epoch Microplate reader, BioTek Instruments). For the gradient table setup, the algal suspension was distributed across 504 wells (42 12-well plates (tissue culture testplates 12 no. 92412, TPP); 2.5 ml of culture per well). Plates were sealed with surgical Micropore tape (3M) to minimize evaporation. The 42 12-well plates were then placed on a table that generates a cross-gradient of temperature (8.6 ± 0.5 °C to 29.2 ± 0.5 °C on the *x* axis) and irradiance (21.0 ± 2.0 to 527.9 ± 14.0 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ on the *y* axis) (Supplementary Table 1). The temperature gradient was generated

using a custom-made table (Labio) equipped with true-daylight LEDs (sTube 2 W 120 ver 11:11, Snaggi) set to a 16 h/8 h light/dark cycle (light from 6 am to 10 pm, Central European winter time). *Mesotaenium* samples exposed to the 504 different conditions for 65 h (for sampling for RNA-seq and physiological measurements) and 89 h (for detailed light microscopy) on the gradient table. Condensed water at the top of the 12-well plates lids was removed three times in the 65 h by lightly tapping the lids twice.

Algal culturing and gradient table setup (pre-experiments n1(I), n1,2(II))

To assess optimal, stress and lethal culture conditions for *Mesotaenium endlicherianum* SAG 12.97 three pre-experiments were performed (n1(I), performed once, and n1,2(II), performed twice). We assessed *Mesotaenium* performance in Woods Hole Medium (WHM)^{82,83} and C-medium⁸¹ for an average of 23.6 days in aerated culture glass flasks (SCHOTT) at 80 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ (12h/12h light/dark cycle (light from 6 am to 6 pm, Central European winter time) at 18 °C). Before the experiment, different cell densities were analysed using a microplate reader and adjusting the culture to $\text{Abs}_{680\text{nm}} 0.06$ (n1(I) or 0.12 (n1,2(II)) (Epoch Microplate reader, BioTek Instruments). For the gradient table setup, the algal suspension was distributed across 504 wells (42 12-well plates (tissue culture testplates 12 no. 92412, TPP); 2.5 ml of culture per well). Plates were sealed with surgical Micropore tape (3M) to minimize evaporation. The 42 12-well plates were then placed on a table that generates a cross-gradient of temperature (for n1(I): 12.7 ± 1.0 °C to 34.0 ± 0.8 °C on the *x* axis; for n1,2(II): 8.6 ± 0.5 °C to 29.2 ± 0.5 °C on the *x* axis) and irradiance (21.0 ± 2.0 to 527.9 ± 14.0 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ on the *y* axis) (Supplementary Table 1 and Supplementary Fig. 1a). The temperature gradient was generated using a custom-made table (Labio) equipped with true-daylight LEDs (sTube 2 W 120 ver 11:11, Snaggi) set to a 16 h/8 h light/dark cycle (light from 6 am to 10 pm, Central European winter time). *Mesotaenium* samples were exposed to the 504 different conditions either for 191 h (n1,2(II)) or 216 h (n1,2(II)) (for absorption spectra measurements). Condensed water at the top of the 12-well plates lids was removed by lightly tapping the lids twice.

Plate reader

In vivo absorbance at 480, 680 and 750 nm of all 42 plates was measured after 65 h exposition (4–6 h after light on) with an absorbance microplate reader Epoch (BioTek Instruments). Nine data points per well were analysed and averaged using Gen5 2.0 software (Biotek), resulting in 108 measurements per 12-well plate per wavelength. For downstream analyses, these values were averaged resulting in one value per 12-well plate per wavelength (Supplementary Fig. 1). After 89 h exposition, 16 plates were chosen from the prominent gradients (the four most extreme conditions in the corners and a cross of vibrant growth along the two gradients) for analysing a full absorption spectrum (300–900 nm) using the same setup (Supplementary Fig. 9 and Supplementary Table 10).

Photophysiological measurements

For maximum quantum yield measurements (F_v/F_m) the maxi version of the IMAGING-PAM (ImagMAX/L, M-series, Walz) with an IMAG-KS CCD camera, controlled with the ImagingWinGigE (V2.32) software, was used. The *Mesotaenium* cultures in the 12-well plates were dark adapted for 10–30 min before measuring. Before measurements, the lid was removed. For the F_v/F_m measurement, a short saturation pulse (intensity 3) was applied. The measurement settings on the IMAGING-PAM were as follows: measuring light 1, gain 3, damping 2 and mean over area of interest was turned off. No special saturation pulse routine was applied to modify the signal-to-noise ratio of the chlorophyll fluorescence measurement.

Statistical analysis of absorption and F_v/F_m values and temperature/light cluster analysis

Statistical analysis of the absorption and the F_v/F_m values was done using a Kruskal–Wallis test with a post hoc Fisher’s least significant difference test⁸⁴ using R (version 4.1.3). *P* values were Bonferroni corrected and grouped into significant groups using R packages ‘agricolae’ version 1.3–5 and ‘dplyr’ version 1.0.9. For heat map generation of physiological values plotted against temperature or light, the R package ‘pheatmap’ version 1.0.12 was used. For cluster analysis, the R package ‘factoextra’ version 1.0.7 was used. Clusters were generated using the *eclust* function with clustering function ‘kmeans’ with the number of clusters set to six and for hierarchical clustering; ‘euclidean’ was used as the distance measure. Clusters were visualized with PCA in R.

RNA extraction and sequencing

After absorption measurements, the 12-well plates were put back on the table to let cells adjust to the table conditions again for a minimum of 5 min before collecting them. For RNA extraction 0.4 ml was taken from every well of the 42 12-well plates on the table after pipetting the cells up and down twice to homogenize them. In total, 4.8 ml liquid culture was taken per condition on the table (that is, pooling 0.4 ml of each 12 wells per each of the 42 conditions). The samples were then centrifuged for 5 min at 20 °C and 4,200g. The supernatant was removed and the pellet was frozen at –80 °C. To extract RNA, the Spectrum Plant Total RNA Kit (STRN250-1KT, Sigma-Aldrich) was used according to the manufacturer’s instructions. For cell disruption, samples in lysis buffer were ultrasonicated for 1 min and vortexed. RNA samples were treated with DNase I (EN0521; Thermo Fisher) and shipped on dry ice to Novogene where they were quality checked with a Bioanalyzer (Agilent Technologies). Libraries were built on the basis of total RNA using poly-T oligo-attached magnetic beads. Following fragmentation, synthesis of the first-strand complementary DNA was carried out using random hexamer primers and second-strand cDNA using dUTP, instead of dTTP. A directional size-selected library was built that included PCR-based amplification. Sequencing adaptors were 5′ adaptor: 5′-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3′ and 3′ adaptor: 5′-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGATGAC TATCTCGTATGCCGTCTTCTGCTTG-3′. The library was sequenced on an Illumina NovaSeq 6000 platform and data were downloaded using *wget* (GNU *Wget* 1.14).

Quality control of reads

We checked the quality of our raw reads via *FastQC*⁸⁴ (v0.11.9) and summarized the results via *MultiQC*⁸⁵ (v1.11). On the basis of these and the used adaptor sequence, we filtered and trimmed reads via *Trimmomatic*⁸⁶ (v0.36) with these parameters: (‘ILLUMINACLIP:novogene_adapter_sequences_Trimmomatic.fa:2:30:10:2:True LEADING:26 TRAILING:26 SLIDINGWINDOW:4:20 MINLEN:36’). We checked the quality of the trimmed reads with *FastQC* and *MultiQC* again.

Genome annotation

The original annotation of *M. endlicherianum*⁸ had a lower number of genes compared with other Zygnematophyceae algae. We took advantage of our newly generated RNA-seq dataset to improve genome annotation. Trimmed reads were mapped via *HISAT2* (ref. 87) and assembled via *StringTie*⁸⁷. The *StringTie* results showed many novel isoforms as well as novel transcripts. We also used *BUSCO* V5 (ref. 88) to measure the completeness of the gene models in annotation V1 independent of *StringTie*. Although the gene prediction method used by *BUSCO* at the genome level is very efficient, it is not unexpected if it misses some proteins that were annotated in a genome via experimental data, based on bioinformatic methods and next-generation sequencing data, or ab initio-based gene prediction methods. Therefore, we expect that the *BUSCO* score based on the proteins of a gene model should be equal to or greater than the *BUSCO* score of the genome. When we compared

the *BUSCO* score between the genome and protein sequences for *M. endlicherianum* with ‘viridiplantae.odb.10-2020-09-10’, we noticed that they show similar numbers (Supplementary Fig. 2). Therefore, we decided to re-annotate the genome of *M. endlicherianum* with our comprehensive RNA-seq datasets as well as public protein and genome sequences published for its close relatives.

We annotated the *M. endlicherianum* genome using *REAT* (v0.6.1). Various gene models were predicted based on different types of evidence and methods. The final gene models and annotation V2 were based on agreement with the experimental evidence. At the end, we tried to quantify ‘completeness’ and quality of the new annotation V2 and the old V1.

First, we used RNA-seq evidence with *REAT*’s ‘Transcriptome Workflow’ with *HISAT2* (v2.2.1), *Scallop*⁸⁹ (v0.10.5) and *StringTie* (v2.1.5). We also used *Portcullis*⁹⁰ (v1.2.4) to identify genuine junctions based on short reads alignments. This workflow uses *Mikado*⁹¹ (v2.3.4) to identify the ‘best’ set of transcripts from multiple transcript assemblies.

Then, we used gene homology information from representative streptophytes in *REAT*’s ‘Homology Workflow’. *SPALN*^{92,93} (v2.4.7) was used to align representative protein sequences onto the *M. endlicherianum* genome. The representative dataset consisted on genome, gene models and protein sequences of *Anthoceros agrestis*⁹⁴ (Oxford strain), *Arabidopsis thaliana*⁹⁵, *Azolla filiculoides*⁹⁶, *Chara braunii*⁷², *Chlorokybus melkonianii*⁹⁷ (for naming, see ref. 98), *Chlamydomonas reinhardtii*⁹⁹ (v5.6), *Klebsormidium nitens*¹⁰⁰, *Mesostigma viride*¹⁰¹, *Marchantia polymorpha*¹⁰² (v6.1r1), *Penium margaritaceum*¹¹, *Physcomitrium patens*¹⁰³ (v3.3), *Selaginella moellendorffii*¹⁰⁴ and *Spiroglaea muscicola*⁸. We also used the junction file produced by *Portcullis*. Since there were no close relatives of *M. endlicherianum* on the *SPALN* species-specific parameter set, we used three different closest possibilities (*Angiosp*, *Chlosp* and *MossWorts*) and built three models. These alignments are filtered using a predefined set of criteria (compare code on GitHub) including exon length, intron length and internal stop codon, among others. The final gene models of V2 were prepared by *Mikado*.

Afterwards, we used *REAT*’s ‘Prediction Workflow’ to predict gene models ab initio and based on RNA-seq and homology evidence. This uses *Augustus*^{105–107} (v3.4.0), *SNAP*¹⁰⁸ (version 2006-07-28), *Glimmer*¹⁰⁹ (v0.3.2) and *CodingQuarry*¹¹⁰ (v2.0), which generate different gene models as the raw material for *EvidenceModeler*¹¹¹ (v1.1.1) that chooses the best set of exons and combine them in a gene model using weights (see GitHub) that could be adjusted for each sort of prediction and evidence. To include untranslated regions where possible, the *EvidenceModeler* output is then processed by *Mikado* using untranslated region-containing gene models from the transcriptome and homology workflows as inputs, as well as gene models classified by *REAT* as gold, silver and bronze based on their agreement with the set of protein sequences from other streptophyte genomes (streptophyte algae and land plants), transcriptome alignment, homology alignment and junctions. To train ab initio predictors, a user-defined number of models are randomly chosen in a user-defined ratio between mono-exonic (10%) and multi-exonic (90%). These models were chosen from best-classified models (gold and silver). For *Augustus*, we performed meta parameter optimization and train a model with *kfold* of 8. Beside ab initio predictions, we used *Augustus* to predict gene models with three different weights for each evidence type as suggested by *REAT* authors (compare code on GitHub).

At last, we used *Minos*¹¹² (v1.8.0), which is gene model consolidation pipeline and produces external metrics based on *DIAMOND*¹¹³ (v0.9.34) ‘BLASTp/BLASTx’, *Kallisto*¹¹⁴ (v0.46.2) expression quantification, coding potential calculator¹¹⁵ (CPC2 v0.1) and *BUSCO* assessments. These metrics pass through *Mikado* in combination with various gene models produced with different methods (as mentioned above); *Minos* determines the best gene model for each region based on user-defined criteria (for details, see GitHub) and external metrics. *Minos* also puts

a tag on each gene model to categorize them based on a user-defined threshold (we used default values) for sequence similarity coverage of homologues, BUSCO score, coding potential calculator score, transcript per million expression and transcript score into 'high confidence', 'low confidence' and 'predicted genes'.

Genome annotation assessment

We used two methods to compare the quality of the new gene model with the published one. We compared the BUSCO scores of the annotated protein sequences as well as genome sequence using the reference 'viridiplantae.odt.10-2020-09-10' dataset. We also used maker¹¹⁶ (v3.01.04) to calculate the AED¹¹⁷ to evaluate the agreement of the gene models with external evidences. Maker-P was used to build the *M. endlicherianum* gene model V1.

Further, we used the maker package to perform functional annotation via InterProScan and BLAST using the agat¹¹⁸ package (v0.9.2). Additionally, we performed a BLAST (v2.11.0+) search against *A. thaliana* protein sequences (Araport11) and reported the best hit for each sequence (Supplementary Table 11) and used eggNOG-mapper^{119,120} (v2.1.8) to perform functional annotation. We used DIAMOND¹¹³ (v2.0.15) with ultra-sensitive mode, with *e* value cut-off of $1e^{-7}$ and in an iterative manner. We used the protein sequences as our inputs and Viridiplantae (33090) as our taxonomy scope.

RNA-seq analysis: pseudoalignment

To quantify gene expression, we used a Snakemake-managed pipeline (7.7.0) that hinged on Kallisto¹¹⁴ (v0.45.0). We indexed the transcriptome file with $-kmer-size=31$ parameter, and used $-bootstrap-samples 100$ and $-rf-stranded$ to quantify gene expression based on pseudo-aligned reads. We used MultiQC to obtain an overview of alignment for each condition.

Filtering, normalization, modelling mean–variance relationship and data exploration

Kallisto quantification files were imported into R (v4.2.0; tidyverse v1.3.1) with tximport¹²¹ (v1.24.0) to calculate the counts from abundance via 'lengthScaledTPM' based on our study design file (Supplementary Table 12). We used edgeR¹²² (v3.38.1) for filtering and trimmed mean of *M*-values normalization¹²³ of the reads (genes with >1 count per million at log₂ scale in at least three samples—the number of replicates—were kept). Then, we used the voom function from limma^{124–127} (v3.52.2) to model the mean–variance relationship. The normalized expression table on the log₂ scale is available in Supplementary Table 13. We performed PCA based on the expression table output of voom and visualized the result with ggplot2 (ref. 128) (v3.3.6). We visualized the heat map of distance and Spearman correlation between all samples considering all genes via pheatmap (v1.0.12), and calculated clusters via the Euclidian method.

RNA-seq analysis: WGCNA

We used the WGCNA^{28,129} package (v1.71) with the expression table produced by limma. We checked for and filtered out outliers as suggested by WGCNA authors (Supplementary Fig. 10). Then, we visualized the scale-free topology model fit (R^2) against the soft thresholds (β) to pick a β for our network construction (Supplementary Fig. 11). We used signed network type and 'bicor' as our correlation function for WGCNA. On the basis of these results, we picked 16 as our soft threshold ' β '. We experimentally chose a merging threshold of 0.25 after exploring different values from 0.2 to 0.4 and investigating the relationship between eigengenes and temperature, light intensity, F_v/F_m and absorption (Supplementary Fig. 12). We built the gene co-expression network using a merging threshold of 0.25 for modules, maximum portion of outliers as 0.05 and minimum module size of 30. Then, we visualized the correlation between each module's eigengene and temperature, light intensity, F_v/F_m and absorption to identify which modules are

more related to each treatment (Fig. 4c). We provided a table for all genes, their module assignment, inter- and intramodular connectivity, gene significance for temperature and light intensity, correlation with temperature and light intensity, and their module membership (that is, signed eigengene-based connectivity) (Supplementary Table 5). We also visualized the graphical representation of the topological overlap matrix (TOM) of our samples (Supplementary Fig. 13). To have a visual representation of gene expression in each module, we drew heat maps for each module via pheatmap (using the Euclidean method for calculating the distance and complete method clustering) (Supplementary Fig. 14). GO enrichment analysis was performed via the clusterProfiler package^{130,131} (v4.4.4) using the output of eggNOG-mapper and adjusted *P* value cut-off of 0.05 and *q* value cut-off of 0.05, considering only genes that are present in our GO term-to-gene table, which was expressed and passed filtering as our background gene universe (Supplementary Table 6). Determining the proper background gene list has major importance in enrichment analysis¹³².

To see how *A. thaliana*'s well-known genes in stress-response mechanisms (downloaded from the TAIR database via keyword search) were distributed across different modules, we performed BLASTp searches against the new *M. endlicherianum* annotated proteins. We visualized the distribution of these IDs for different stress-related keywords (Supplementary Fig. 15) and the expression of these genes across different samples via pheatmap (Supplementary Fig. 16). We defined as module hubs the top 20 genes (nodes) with the highest connectivity within each module (Supplementary Tables 5 and 14).

Differential gene expression analysis

We performed differential gene expression analysis using the limma package. We divided samples into multiple groups as follows: low light intensity (21 and 39 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), medium light intensity (72 and 129 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), high light intensity (329 and 527 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), low temperature (8 and 12 °C), medium temperature (17, 20 and 23 °C) and high temperature (26 and 29 °C; see grid/coloured table layout in Fig. 2). We performed all-against-all comparisons and an additional comparison of those samples from an $F_w/F_m < 0.5$ versus low light intensity + medium temperature. We used duplicateCorrelation as suggested by Smyth et al.¹³³ to consider technical replicates. We used clusterProfiler for GO enrichment analysis¹³¹ with an adjusted *P* value and *q* value cut-off of 0.01 and only genes that were expressed and passed filtering as our background universe. The heat map of gene expression profiles, dot plot and cnetplot of enriched GO terms for each comparison is available in Supplementary Table 14 and Supplementary Figs. 17–25).

Phylogenetic analyses

We assembled a protein database based on the protein releases from the genomes of: *Anthoceros agrestis* BONN⁹⁴, *Anthoceros punctatus*⁹⁴, *Amborella trichopoda*¹³⁴, *Arabidopsis thaliana*¹³⁵, *Azolla filiculoides*⁹⁶, *Bathycoccus prasinos*¹³⁶, *Brassica oleracea*¹³⁷, *Brassica rapa*¹³⁸, *Brachypodium distachyon*¹³⁹, *Capsella grandiflora*¹⁴⁰, *Chara braunii*⁷², *Chlorokybus melkonianii*⁹⁷ (for naming, see ref. 98), *Chlamydomonas reinhardtii*⁹⁹, *Coccomyxa subellipsoidea*¹⁴¹, *Gnetum montanum*¹⁴², *Klebsormidium nitens*¹⁰⁰, *Marchantia polymorpha*¹⁴³, *Mesostigma viride*⁹⁷, *Micromonas pusilla*¹⁴⁴, *Micromonas* sp.¹⁴⁴, *Oryza sativa*¹⁴⁵, *Picea abies*¹⁴⁶, *Physcomitrium patens*¹⁰³, *Salvinia cucullata*⁹⁶, *Selaginella moellendorffii*¹⁰⁴, *Solanum lycopersicum*¹⁴⁷, *Theobroma cacao*¹⁴⁸, *Mesotaenium endlicherianum*⁸, *Ostreococcus lucimarinus*¹⁴⁹, *Penium margaritaceum*¹¹, *Spiroglora muscicola*⁸, *Ulva mutabilis*¹⁵⁰, *Volvox carteri*¹⁵¹, *Isoetes taiwanensis*¹⁵² and *Ceratopteris richardii*¹⁵³.

Homologues for proteins were detected using BLASTp with *Arabidopsis* and *Mesotaenium* proteins as query against the aforementioned proteins as database. Alignments were computed using MAFFT v7.490 (ref. 154). All phylogenies were computed with IQ-TREE¹⁵⁵ multicore version 1.5.5; their respective best-fit model for protein evolution was

determined using ModelFinder¹⁵⁶ (integrated in IQ-TREE multicore version 1.5.5 for Linux 64-bit built 2 June 2017) according to Bayesian Information Criterion; and 1,000 ultrafast bootstrap¹⁵⁷ pseudoreplicates were carried out and 100 non-parametric bootstrap¹⁵⁸ pseudoreplicates for the LDAP phylogeny. We coloured phylogeny trees via ggtree (v3.9.0).

DIC and confocal laser scanning microscopy

Differential interference contrast (DIC) imaging was done for all replicates from the table with an Olympus BX-60 microscope (Olympus, Japan) with a ProgRes C14plus camera and the ProgRes CapturePro Software (version 2.9.01) (JENOPTIK AG). The morphology of chosen conditions (Fig. 1, Extended Data Figs. 4–6 and Supplementary Fig. 1) of *Mesotaenium* cells that were 89 h on the table was analysed.

For algae that were used for quantifying the abundance of LD per cell, a ZEISS AxioScope 7 microscope (Carl Zeiss) was used including the Zen software (Carl Zeiss). The LD count was carried out in Fiji¹⁵⁹. For statistical analysis of the LD count data, we first used a Shapiro–Wilk test¹⁶⁰ to assess normality and used Mann–Whitney *U* tests¹⁶¹ with R (version 3.6.1) accordingly.

Confocal laser scanning microscope was done on a Zeiss LSM780 (Carl Zeiss) set as in Müller et al.¹⁶². For the staining of the LD structures, we used the neutral lipid specific stain BODIPY 493/503 (EM/EX) (Merck). *Mesotaenium* cells were grown for 22 days on WHM medium at 70–80 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ and 22 °C. These cells were ultrasonicated for 1 min with 1:500 BODIPY and incubated on a shaker for 5 min before visualization.

LD isolation and proteomics

For LD isolation 23-day-old *Mesotaenium* cells grown on WHM medium at 70–80 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ and 22 °C were homogenized using a Tenbroeck or potter homogenizer in LD isolation buffer (10 mM sodium phosphate buffer pH 7.5, 200 μM phenylmethylsulfonyl fluoride, 0.5 mM dithiobis(succinimidyl propionate) and 10 mM *N*-ethylmaleimide). The resulting centrifuged supernatant of a 100g spin for 1 min was considered as TE. After two further high-speed centrifugations (SW40 Ti for 1 h, 4 °C at 100,000g, TLA120 for 1 h at 100,000g and 4 °C) the floating fat pad was precipitated at –20 °C using 100% ethanol overnight. The precipitated pellet was washed with 80% ethanol twice, dried and then suspended in 6 M urea. Protein concentration was determined using the bicinchoninic acid assay. An in-gel sodium dodecyl sulphate gel digestion was done with trypsin adapted from Shevchenko et al.¹⁶³. C18 Stage tip purification was done according to Rappsilber et al.^{164,165}. Protein samples were analysed using LC–MS. For this, peptide samples were reconstituted in 20 μl LC–MS sample buffer (2% acetonitrile and 0.1% formic acid). Then, 2 μl of each sample were subjected to reverse-phase liquid chromatography for peptide separation using an RSLCnano UltiMate 3000 system (Thermo Fisher Scientific). Peptides were loaded on an Acclaim PepMap 100 pre-column (100 $\mu\text{m} \times 2 \text{ cm}$, C18, 5 μm , 100 Å; Thermo Fisher Scientific) with 0.07% trifluoroacetic acid at a flow rate of 20 $\mu\text{l min}^{-1}$ for 3 min. Analytical separation of peptides was done on an Acclaim PepMap RSLC column (75 $\mu\text{m} \times 50 \text{ cm}$, C18, 2 μm , 100 Å; Thermo Fisher Scientific) at a flow rate of 300 nl min^{-1} . The solvent composition was gradually changed within 94 min from 96% solvent A (0.1% formic acid) and 4% solvent B (80% acetonitrile and 0.1% formic acid) to 10% solvent B within 2 min, to 30% solvent B within the next 58 min, to 45% solvent B within the following 22 min and to 90% solvent B within the last 12 min of the gradient. All solvents and acids had Optima grade for LC–MS (Thermo Fisher Scientific). Eluting peptides were on-line ionized by nano-electrospray using the Nanospray Flex Ion Source (Thermo Fisher Scientific) at 1.5 kV (liquid junction) and transferred into a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Full scans in a mass range of 300–1,650 m/z were recorded at a resolution of 30,000 followed by data-dependent top ten higher energy collisional

dissociation fragmentation at a resolution of 15,000 (dynamic exclusion enabled). LC–MS method programming and data acquisition was performed with the XCalibur 4.0 software (Thermo Fisher Scientific). Afterwards, the raw proteome data were analysed using Max Quant software¹⁶⁶ version 1.6.2.10. The database for this analysis was our new V2 gene model data. The data were then further processed by the Perseus (1.6.2.2) software^{166,167}.

Lipid analysis of LDs

LDs (200–300 μl) were extracted with 10 ml of methanol/chloroform/formic acid (20:10:1, vol/vol/vol), 5 ml of 0.2 M phosphoric acid and 1 M potassium chloride¹⁶⁸. After vortexing and centrifugation at 50g for 2 min, the lower chloroform phases were dried under streaming nitrogen and redissolved in chloroform/methanol (2:1, vol/vol). For analytical analysis, one-fifth of the lipid extracts were spotted on a thin layer chromatography (TLC) silica plate (TLC Silica gel 60, 20 \times 20 cm, Merck KGaG) and separated with petroleum ether/diethyl ether/acetic acid (70:30:0.5, vol/vol/vol)¹⁶⁹. The lipid composition was identified after incubation in copper sulphate solution (0.4 M CuSO_4 in 6.8% (vol/vol) phosphoric acid) and heating at 180 °C. For preparative analysis, half of the lipid extracts were additionally separated by TLC. After development, the lipid spots were visualized after spraying with 0.05% (wt/vol) primuline in 80% (vol/vol) acetone. The silica gel spots containing TAG, diacylglycerol (DAG) and free fatty acids (FFA) were used for preparation of fatty acid methyl esters as already described¹⁷⁰ with some modifications. For acidic hydrolysis, 1 ml of methanol/toluene (2:1, vol/vol) containing 2.75% (vol/vol) sulphuric acid (95–97%) and 2% (vol/vol) dimethoxypropane was added to the scraped silica gel. For quantification, 1 μg of tripentadecanoate was added and the samples were incubated for 1 h at 80 °C. To extract the resulting fatty acid methyl esters, 1.5 ml of saturated aqueous sodium chloride solution and 1.2 ml of hexane were added and centrifuged at 450g for 10 min. The hexane phase was dried under streaming nitrogen and redissolved in 10 μl acetonitrile. Gas chromatography (GC) analysis was performed with an Agilent (Waldbronn, Germany) 6890 gas chromatograph fitted with a capillary DB-23 column (30 $\text{m} \times 0.25 \text{ mm}$; 0.25 μm coating thickness; J&W Scientific, Agilent) modified from Hornung et al.¹⁷¹. Helium was used as carrier gas at a flow rate of 1 ml min^{-1} . The temperature gradient was 150 °C for 1 min, 150–200 °C at 4 K min^{-1} , 200–250 °C at 5 K min^{-1} and 250 °C for 6 min. The peak area was collected with the ChemStation software (Agilent). From the absolute fatty acid contents, relative fatty acid profiles for TAG, DAG and FFA were calculated.

Pollen tube transformation and microscopy

Coding sequences for *Mesotaenium* homologues salient to LD biology were *MeCaleosinf* 5'-GGGGACAAGTTTG TACAAAAAAGCAGGCTCATGTCGAAGCTCAGTCTTGCC-3', *MeCaleosinr* 5'-GGGGACCCTTTGTACAAGAAAGCTGGGTCA GACTGCTTCTTCTCTGCTT-3', *MeLDAPf* 5'-GGGGACAAGTTT GTACAAAAAAGCAGGCTCATGGCCGAAAGTCAGGGCCC-3', *MeLDAPr* 5'-GGGGACCCTTTGTACAAGAAAGCTGGGTCC GACTTCTTGAGGGCGTCCGGC-3', *MeSteroleosinf* 5'-GGGGACAA GTTTGTACAAAAAAGCAGGCTCATGGGTACTTAATGCCCTTGC-3', *MeSterleosinr* 5'-GGGGACCCTTTGTACAAGAAAGCTGGGTCCG CATTGGACTTGACGAGGG-3', *MeOleosinf* 5'-GGGGACAAGTTTG TACAAAAAAGCAGGCTCATGCCTCAGGATCAGCAGCAAG-3', and *MeO leosinr* 5'-GGGGACCCTTTGTACAAGAAAGCTGGGTCTTCTCTC CTCTCAACCTTGT-3'. Constructs for expression in pollen tubes were cloned into the pLatMCC-GW vector using the fast Gateway method as described previously¹⁶². Pollen transformation, pollen tube growth and fixation were also performed according to this protocol. LDs were stained with BODIPY 493/503 at a final concentration of 1.3 $\mu\text{g ml}^{-1}$. Microscopy images of transformed tobacco pollen tubes were acquired using an LSM 980 confocal laser scanning microscope using the objective C-Apochromat 40 \times /1.20 W Korr (both Carl Zeiss). mCherry was

excited at 561 nm and detected at 600–640 nm. BODIPY 493/503 was excited at 488 nm and detected at 490–535 nm. In both cases, the major beam splitter MBS 488/561 was used. Both channels were recorded independently using the line mode.

Comparative evolutionary analyses

To perform comparative evolutionary analyses among *M. endlicherianum*, other streptophyte algae and embryophytes, we used two separate workflows based on one criterion: the availability of at least 15 raw RNA-seq samples for a given species challenged with abiotic stresses and control conditions. This is the minimum requirement to build a co-expression network using the WGCNA package. If a species passed this criterion, we used them in two approaches; all results from this comparative analyses can be found in Supplementary Table 15.

Approach 1: to compare co-expression networks computed based on control and abiotic stress samples, we first used Orthofinder and protein sequences of *A. agrestis*⁹⁴, *A. filiculoides*⁹⁶, *A. thaliana*¹³⁵, *B. distachyon*¹³⁹, *C. braunii*⁷², *Closterium* sp. NIES 67 (ref. 10), *C. melkonianii*⁹⁷, *C. reinhardtii*⁹⁹, *K. nitens*¹⁰⁰, *M. endlicherianum*⁸, *M. polymorpha*¹⁰², *M. viride*⁹⁷, *O. sativa*¹⁴⁵, *P. margaritaceum*¹¹, *P. patens*¹⁰³, *S. lycopersicum*¹⁴⁷, *S. moellendorffii*¹⁰⁴, *S. muscicola*⁸, *Z. mays*¹⁷² and *Zygnema circumcarinatum*⁹ as well as a species cladogram to find phylogenetic HOGs using these parameters: -S mmseqs -M msa -A mafft -s species_tree.txt -y. For *A. thaliana*, we downloaded a gene-GO table from arabis.dopsis.org. For *P. patens*, *M. polymorpha* and *Zygnema 1b*, we used eggNOG-mapper and their protein sequences to create a gene-GO table using these parameters: -m diamond -dmnd_iterate yes -eval 1e-10 -sensmode ultra-sensitive -tax_scope 33090. We downloaded raw RNA-seq reads for *A. thaliana*¹⁷³⁻¹⁷⁷, *P. patens*¹⁷⁸ and accessions PRJNA277025 and PRJNA192876, *M. polymorpha*¹⁷⁹⁻¹⁸¹ and *Zygnema 1b* (ref. 9) from the National Center for Biotechnology Information (NCBI). We followed the same quantification as *Mesotaenium* for each species here. In short, we used FastQC, MultiQC and Trimmomatic to check the quality of each read and filter and trim the raw reads. Then, we used Kallisto to pseudoalign the reads to the transcriptome of that species. Then, we imported gene counts for each species into R and performed similar exploratory analyses to *Mesotaenium* for each species. An additional layer of analysis here was to check for batch effect when we looked at all samples from different sources for a species. We used hierarchical clustering and PCA to pick the best expression profile from (i) uncorrected, (ii) batch-corrected as a covariate using limma, and (iii) batch-corrected using ComBat-seq¹⁸² to adjust for batch effects (if there were any). There is a debate in the community about which method is the best practice; therefore, we did all for every species and picked the best (less confounding effect between batches and maximum similarity between similar conditions) for each species. Then, we used the expression profile and built a signed co-expression network using the WGCNA package for each species. We followed the same procedure as *Mesotaenium*. We performed a GO enrichment analysis for each module in the co-expression networks. Then, we used the Orthofinder-based orthogroups to find genes that have a counterpart in *Mesotaenium* for each species and then we calculated the Jaccard similarity and dissimilarity between each *Mesotaenium* modules and each module of *A. thaliana*, *P. patens*, *M. polymorpha* and *Zygnema circumcarinatum* SAG698-1b. For each module in these co-expression networks, we looked for the connectivity of genes that share a HOG with *Mesotaenium* hubs.

Approach 2: to determine the shared DEGs under abiotic stresses across streptophyte algae, we first downloaded raw RNA-seq reads from NCBI as follows: (1) *Mougeotia*^{24,25} sp. MZCH240 and *S. pratensis* MZCH10213 (ref. 24), (2) *M. viride*, *C. cerffii*, *K. flaccidum*, *C. globularis*, *C. scutata*, *Zygnema cylindricum*²⁰ SAG698-1a, (3) *Zygnema* sp.²³ SAG2419, (4) *S. pratensis*²² UTEX92 and (5) *Z. circumcarinatum*⁹. If it was possible, we also obtained the transcriptome or genome file for each species. Then, we used Orthofinder and protein sequences of

Mesotaenium and the protein sequences of these species as well as a species cladogram to find phylogenetic hierarchical orthogroups using these parameters: -S mmseqs -M msa -A mafft -s species_tree.txt -y. For species for which only the transcriptome was available, we used TransDecoder (v5.7.0) using TransDecoder.LongOrfs and TransDecoder.Predict scripts to get a protein-coding sequence for our Orthofinder run. For those that we did not have a transcriptome, we built one using Trinity (v2.15.1) (ref. 183) and the settings -seqType fq -trimmomatic. We followed the same quantification steps as for *Mesotaenium* and workflow A to pseudoalign reads to the transcriptome. Then, we followed similar steps to *Mesotaenium* to calculate DEGs for each species. Finally, we compared these DEGs with DEGs in *Mesotaenium* using HOGs from Orthofinder run in this workflow. We used BioNERO package¹⁸⁴ to aggregate log₂ (fold change) values for each gene in each species to the corresponding HOGs and then used cluster Profiler to perform GO enrichment analyses and visualized the heat maps. In all comparisons, we considered adjusted *P* values <0.05 as significant enrichment.

TEs

We used InterProScan¹⁸⁵ (v5.59-91.0) on all predicted proteins in *Mesotaenium endlicherianum* V2 and filtered the results for transposon-related domains. This resulted in 6,186 entries in 1,748 unique gene IDs, among which only 96 were expressed in our RNA-seq data (that is, passing an expression cut-off of at least 1 count per million in at least three samples); all results are presented in Supplementary Table 4b).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All RNA-seq reads have been uploaded to the NCBI Sequence Read Archive and can be accessed under BioProject PRJNA832564 and Sequence Read Archive accessions SRR18936040 to SRR18936170. Furthermore, data can be interactively explored at <https://mesotaenium.uni-goettingen.de> and proteomic data have been uploaded to EMBL-EBI PRIDE (accession PXD037847). On Zenodo, we have deposited (1) raw light and confocal micrographs generated, for example, for LD assessment in *Mesotaenium* and pollen tubes <https://doi.org/10.5281/zenodo.7921367> and (2) raw and visualized phylogenetic data <https://doi.org/10.5281/zenodo.7950653>. The additional previously published RNA-seq datasets that were used for comparisons are: (1) *A. thaliana*: SRR2302908 to SRR2302919, ERR754084, ERR754066, ERR754077, ERR754069, ERR754087, ERR754064, ERR754059, SRR7659142, SRR7659143, SRR7659144, SRR7659145 to SRR7659150, SRR5197904, to SRR5197909; (2) *M. polymorpha*: SRR12076853, SRR12076855, SRR12076857, SRR12076859, SRR12076861, SRR12076863, SRR12076865, SRR12076867, SRR12076869, SRR12076871, SRR12076873, SRR12076875, SRR12076877, SRR12076879, SRR12076917 to SRR12076925, SRR15186078 to SRR15186125, DRR093991 to DRR093996; (3) *P. patens*: SRR1824306 to SRR1824320, SRR10235460 to SRR10235483, SRR787291, SRR787292, SRR787293, SRR787294, SRR787295; (4) *Z. circumcarinatum* SAG698-1b: SRR24939299, SRR24940177, SRR24909175, SRR24757807, SRR24757829, SRR24757830, SRR24757831, SRR24205691 to SRR24205702, SRR24286545 to SRR24286562, SRR24576622, SRR24576623, SRR24385702, SRR24450996, SRR24450997, SRR24451196, SRR24480449, SRR24707416, SRR24707417, SRR24952091, SRR21891679 to SRR21891705; (5) *C. cerffii* (at the time, *C. atmophyticus*, see ref. 98): SRR5949009, SRR5949013 to SRR5949016, SRR5949027 to SRR5949030; (6) *C. scutata*: SRR5948993, SRR5948995 to SRR5948998, SRR5949001, SRR5949004, SRR5949005, SRR5949007; (7) *K. flaccidum*: SRR5949010, SRR5949011, SRR5949012, SRR5990072 to

SRR5990080; (8) *M. viride*: SRR5949021 to SRR5949026; (9) *Mougeotia* sp. MZCH240: SRR9083681, SRR9083682, SRR9083688, SRR9083692 to SRR9083701; (10) *S. pratensis* MZCH10213: SRR9083685, SRR9083686, SRR9083687, SRR9083689, SRR9083690, SRR9083696; (11) *S. pratensis* UTEX928: SRR4018077 to SRR4018100; (12) *Z. circumcarinatum* SAG698-1a: SRR5948999, SRR5949000, SRR5949002, SRR5949003, SRR5949006, SRR5949008, SRR5949017, SRR5949018; and (13) *Z. circumcarinatum* SAG2419: SRR6047298, SRR6047299, SRR6047302 to SRR6047305. Source data are provided with this paper.

Code availability

Codes and data used for genome re-annotation, WGCNA and differential gene expression analysis are available on our GitHub page https://github.com/deVries-lab/Response_to_a_gradient_of_environmental_cues_in_mesotaenium_endlicherianum.

References

- Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl Acad. Sci. USA* **115**, 6506–6511 (2018).
- Lenton, T. M. et al. Earliest land plants created modern levels of atmospheric oxygen. *Proc. Natl Acad. Sci. USA* **113**, 9704–9709 (2016).
- Wodniok, S. et al. Origin of land plants: do conjugating green algae hold the key? *BMC Evol. Biol.* **11**, 104 (2011).
- Wickett, N. J. et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868 (2014).
- Puttick, M. N. et al. The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr. Biol.* **28**, 733–745 (2018).
- One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
- Hess, S. et al. A phylogenomically informed five-order system for the closest relatives of land plants. *Curr. Biol.* **32**, 4473–4482 (2022).
- Cheng, S. et al. Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**, 1057–1067.e14 (2019).
- Feng, X. et al. Chromosome-level genomes of multicellular algal sisters to land plants illuminate signaling network evolution. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.01.31.526407> (2023).
- Sekimoto, H. et al. A divergent RWP-RK transcription factor determines mating type in heterothallic *Closterium*. *N. Phytol.* <https://doi.org/10.1111/nph.18662> (2023).
- Jiao, C. et al. The *Penium margaritaceum* genome: hallmarks of the origins of land plants. *Cell* **181**, 1097–1111.e12 (2020).
- Golicz, A. A., Batley, J. & Edwards, D. Towards plant pangenomics. *Plant Biotechnol. J.* **14**, 1099–1105 (2016).
- Gordon, S. P. et al. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* **8**, 2184 (2017).
- Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J. & Edwards, D. Plant pan-genomes are the new reference. *Nat. Plants* **6**, 914–920 (2020).
- Umezawa, T. et al. Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant Cell Physiol.* **51**, 1821–1839 (2010).
- Bowman, J. L., Briginshaw, L. N., Fisher, T. J. & Flores-Sandoval, E. Something ancient and something neofunctionalized—evolution of land plant hormone signaling pathways. *Curr. Opin. Plant Biol.* **47**, 64–72 (2019).
- Hundertmark, M. & Hincha, D. K. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118–122 (2008).
- Carella, P. et al. Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Curr. Biol.* **29**, 2282–2294.e5 (2019).
- Rieseberg, T. P. et al. Crossroads in the evolution of plant specialized metabolism. *Sem. Cell Dev. Biol.* **134**, 37–58 (2023).
- de Vries, J., Curtis, B. A., Gould, S. B. & Archibald, J. M. Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc. Natl Acad. Sci. USA* **115**, E3471–E3480 (2018).
- Sun, Y. et al. A ligand-independent origin of abscisic acid perception. *Proc. Natl Acad. Sci. USA* **116**, 24892–24899 (2019).
- Van de Poel, B., Cooper, E. D., Van Der Straeten, D., Chang, C. & Delwiche, C. F. Transcriptome profiling of the green alga *Spirogyra pratensis* (Charophyta) suggests an ancestral role for ethylene in cell wall metabolism, photosynthesis, and abiotic stress responses. *Plant Physiol.* **172**, 533–545 (2016).
- Rippin, M., Becker, B. & Holzinger, A. Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant Cell Physiol.* **58**, 2067–2084 (2017).
- de Vries, J. et al. Heat stress response in the closest algal relatives of land plants reveals conserved stress signaling circuits. *Plant J.* **103**, 1025–1048 (2020).
- Fürst-Jansen, J. M. R. et al. Submergence of the filamentous Zygnematophyceae *Mougeotia* induces differential gene expression patterns associated with core metabolism and photosynthesis. *Protoplasma* **259**, 1157–1174 (2022).
- Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
- Edel, K. H., Marchadier, E., Brownlee, C., Kudla, J. & Hetherington, A. M. The evolution of calcium-based signalling in plants. *Curr. Biol.* **27**, R667–R679 (2017).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
- Song, J.-Y., Leung, T., Ehler, L. K., Wang, C. & Liu, Z. Regulation of meristem organization and cell division by *TSO1*, an *Arabidopsis* gene with cysteine-rich repeats. *Development* **127**, 2207–2217 (2000).
- Kleine, T. et al. Acclimation in plants—the Green Hub consortium. *Plant J.* **106**, 23–40 (2021).
- Moreno, J. C., Mi, J., Alagoz, Y. & Al-Babili, S. Plant apocarotenoids: from retrograde signaling to interspecific communication. *Plant J.* **105**, 351–375 (2021).
- Rossini, L., Cribb, L., Martin, D. J. & Langdale, J. A. The maize Golden2 gene defines a novel class of transcriptional regulators in plants. *Plant Cell* **13**, 1231–1244 (2001).
- Yasumura, Y., Moylan, E. C. & Langdale, J. A. A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell* **17**, 1894–1907 (2005).
- Waters, M. T. et al. GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell* **21**, 1109–1128 (2009).
- Timm, S. et al. A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in *Arabidopsis*. *Plant Cell* **20**, 2848–2859 (2008).
- Fürst-Jansen, J. M. R., de Vries, S. & de Vries, J. Evo-physio: on stress responses and the earliest land plants. *J. Exp. Bot.* **71**, 3254–3269 (2020).
- Liu, L.-J. et al. COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in *Arabidopsis*. *Plant Cell* **20**, 292–306 (2008).
- Sarid-Krebs, L. et al. Phosphorylation of CONSTANS and its COP1-dependent degradation during photoperiodic flowering of *Arabidopsis*. *Plant J.* **84**, 451–463 (2015).

39. Ordoñez-Herrera, N. et al. The transcription factor COL12 is a substrate of the COP1/SPA E3 ligase and regulates flowering time and plant architecture. *Plant Physiol.* **176**, 1327–1340 (2018).
40. Dai, S. et al. BROTHER OF LUX ARRHYTHMO is a component of the *Arabidopsis* circadian clock. *Plant Cell* **23**, 961–972 (2011).
41. Ju, C. et al. Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nat. Plants* **1**, 14004 (2015).
42. Kato, Y. & Sakamoto, W. Protein quality control in chloroplasts: a current model of D1 protein degradation in the photosystem II repair cycle. *J. Biochem.* **146**, 463–469 (2009).
43. Kato, Y., Sun, X., Zhang, L. & Sakamoto, W. Cooperative D1 degradation in the photosystem II repair mediated by chloroplastic proteases in *Arabidopsis*. *Plant Physiol.* **159**, 1428–1439 (2012).
44. Sjögren, L. L. E., Stanne, T. M., Zheng, B., Sutinen, S. & Clarke, A. K. Structural and functional insights into the chloroplast ATP-dependent Clp protease in *Arabidopsis*. *Plant Cell* **18**, 2635–2649 (2006).
45. Nishimura, K., Kato, Y. & Sakamoto, W. Chloroplast proteases: updates on proteolysis within and across suborganellar compartments. *Plant Physiol.* **171**, 2280–2293 (2016).
46. Pfalz, J., Liere, K., Kandlbinder, A., Dietz, K.-J. & Oelmüller, R. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* **18**, 176–197 (2006).
47. Susek, R. E., Ausubel, F. M. & Chory, J. Signal transduction mutants of *Arabidopsis* uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell* **74**, 787–799 (1993).
48. Jiao, Y., Lau, O. S. & Deng, X. W. Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genet.* **8**, 217–230 (2007).
49. Chen, R. E. & Thorner, J. Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta.* **1773**, 1311–1340 (2007).
50. Nakagami, H., Pitzschke, A. & Hirt, H. Emerging MAP kinase pathways in plant stress signalling. *Trends Plant Sci.* **10**, 339–346 (2005).
51. Rodriguez, M. C. S., Petersen, M. & Mundy, J. Mitogen-activated protein kinase signaling in plants. *Annu. Rev. Plant Biol.* **61**, 621–649 (2010).
52. Meng, X. & Zhang, S. MAPK cascades in plant disease resistance signaling. *Annu. Rev. Phytopathol.* **51**, 245–266 (2013).
53. Chen, X. et al. Protein kinases in plant responses to drought, salt, and cold stress. *J. Integr. Plant Biol.* **63**, 53–78 (2021).
54. Sun, Y. et al. Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. *Dev. Cell* **19**, 765–777 (2010).
55. Planas-Riverola, A. et al. Brassinosteroid signaling in plant development and adaptation to stress. *Development* **146**, dev151894 (2019).
56. Hématy, K. et al. A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17**, 922–931 (2007).
57. Schröder, F., Lisso, J., Lange, P. & Müssig, C. The extracellular EXO protein mediates cell expansion in *Arabidopsis* leaves. *BMC Plant Biol.* **9**, 20 (2009).
58. Schindelman, G. et al. COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in *Arabidopsis*. *Genes Dev.* **15**, 1115–1127 (2001).
59. Roudier, F., Schindelman, G., DeSalle, R. & Benfey, P. N. The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion. *Plant Physiol.* **130**, 538–548 (2002).
60. Ko, J.-H., Kim, J. H., Jayanty, S. S., Howe, G. A. & Han, K.-H. Loss of function of COBRA, a determinant of oriented cell expansion, invokes cellular defence responses in *Arabidopsis thaliana*. *J. Exp. Bot.* **57**, 2923–2936 (2006).
61. Wolf, S. et al. A receptor-like protein mediates the response to pectin modification by activating brassinosteroid signaling. *Proc. Natl Acad. Sci. USA* **111**, 15261–15266 (2014).
62. Higashi, Y., Okazaki, Y., Myouga, F., Shinozaki, K. & Saito, K. Landscape of the lipidome and transcriptome under heat stress in *Arabidopsis thaliana*. *Sci. Rep.* **5**, 10533 (2015).
63. Mueller, S. P., Krause, D. M., Mueller, M. J. & Fekete, A. Accumulation of extra-chloroplastic triacylglycerols in *Arabidopsis* seedlings during heat acclimation. *J. Exp. Bot.* **66**, 4517–4526 (2015).
64. Gidda, S. K. et al. Lipid droplet-associated proteins (LDAPs) are required for the dynamic regulation of neutral lipid compartmentation in plant cells. *Plant Physiol.* **170**, 2052–2071 (2016).
65. Doner, N. M. et al. *Arabidopsis thaliana* EARLY RESPONSIVE TO DEHYDRATION 7 localizes to lipid droplets via its senescence domain. *Front. Plant Sci.* **12**, 658961 (2021).
66. Krawczyk, H. E. et al. Heat stress leads to rapid lipid remodeling and transcriptional adaptations in *Nicotiana tabacum* pollen tubes. *Plant Physiol.* **189**, 490–515 (2022).
67. Listenberger, L. L. & Brown, D. A. Fluorescent detection of lipid droplets and associated proteins. *Curr. Protoc. Cell Biol.* **35**, 24.2.1–24.2.11 (2007).
68. Kretzschmar, F. K. et al. Identification of low-abundance lipid droplet proteins in seeds and seedlings. *Plant Physiol.* **182**, 1326–1345 (2020).
69. Lass, A. et al. Adipose triglyceride lipase-mediated lipolysis of cellular fat stores is activated by CGI-58 and defective in Chanarin–Dorfman syndrome. *Cell Metab.* **3**, 309–319 (2006).
70. James, C. N. et al. Disruption of the *Arabidopsis* CGI-58 homologue produces Chanarin–Dorfman-like lipid droplet accumulation in plants. *Proc. Natl Acad. Sci. USA* **107**, 17833–17838 (2010).
71. Guzha, A., Whitehead, P., Ischebeck, T. & Chapman, K. D. Lipid droplets: packing hydrophobic molecules within the aqueous cytoplasm. *Annu. Rev. Plant Biol.* **74**, 195–223 (2023).
72. Nishiyama, T. et al. The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464.e24 (2018).
73. Zhao, C. et al. Evolution of chloroplast retrograde signaling facilitates green plant adaptation to land. *Proc. Natl Acad. Sci. USA* **116**, 5015–5020 (2019).
74. Honkanen, S. & Small, I. The GENOMES UNCOUPLED1 protein has an ancient, highly conserved role but not in retrograde signalling. *New Phytol.* **236**, 99–113 (2022).
75. Martín, G. et al. Phytochrome and retrograde signalling pathways converge to antagonistically regulate a light-induced transcriptional network. *Nat. Commun.* **7**, 11431 (2016).
76. Gasulla, F. et al. The response of *Asterochloris erici* (Ahmadjian) Skaloud et Peksa to desiccation: a proteomic approach. *Plant Cell Environ.* **36**, 1363–1378 (2013).
77. Li-Beisson, Y., Thelen, J. J., Fedosejevs, E. & Harwood, J. L. The lipid biochemistry of eukaryotic algae. *Prog. Lipid Res.* **74**, 31–68 (2019).
78. de Vries, J. & Ischebeck, T. Ties between stress and lipid droplets pre-date seeds. *Trends Plant Sci.* **25**, 1203–1214 (2020).
79. *The Culture Collection of Algae at the University of Göttingen, Germany (SAG)* (The University of Göttingen); https://sagdb.uni-goettingen.de/detailedList.php?str_number=12.97
80. Friedl, T. & Lorenz, M. The Culture Collection of Algae at Göttingen University (SAG): a biological resource for biotechnological and biodiversity research. *Procedia Environ. Sci.* **15**, 110–117 (2012).
81. Ichimura, T. Sexual cell division and conjugation-papilla formation in sexual reproduction of *Closterium strigosum*. In *Proc. 7th International Seaweed Symposium* 208–214 (Univ. of Tokyo Press, 1971).

82. Nichols, H. W. in *Handbook of Phycological Methods* (ed. Stein J. R.) p. 16–17 (Cambridge Univ. Press, 1973).
83. Conover, W. J. *Practical Nonparametric Statistics* 3rd edn (John Wiley & Sons, 1999).
84. Simon, A. FastQC: a quality control tool for high throughput sequence data (Babraham Bioinformatics, Babraham Institute, 2010); <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
85. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
86. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
87. Perteza, M., Kim, D., Perteza, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
88. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
89. Shao, M. & Kingsford, C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat. Biotechnol.* **35**, 1167–1169 (2017).
90. Mapleson, D., Venturini, L., Kaithakottil, G. & Swarbreck, D. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience* **7**, giy131 (2018).
91. Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience* **7**, giy093 (2018).
92. Gotoh, O. Direct mapping and alignment of protein sequences onto genomic sequence. *Bioinformatics* **24**, 2438–2444 (2008).
93. Gotoh, O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res.* **36**, 2630–2638 (2008).
94. Li, F.-W. et al. *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6**, 259–272 (2020).
95. Cheng, C.-Y. et al. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
96. Li, F.-W. et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **4**, 460–472 (2018).
97. Wang, S. et al. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* **6**, 95–106 (2020).
98. Irisarri, I. et al. Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc. Biol. Sci.* **288**, 20212168 (2021).
99. Merchant, S. S. et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).
100. Hori, K. et al. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* **5**, 3978 (2014).
101. Liang, Z. et al. *Mesostigma viride* genome and transcriptome provide insights into the origin and evolution of Streptophyta. *Adv. Sci.* **7**, 1901850 (2019).
102. Montgomery, S. A. et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).
103. Lang, D. et al. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).
104. Banks, J. A. et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* **332**, 960–963 (2011).
105. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
106. Stanke, M., Tzvetkova, A. & Morgenstern, B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7**, S11 (2006).
107. Hoff, K. J. & Stanke, M. Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinformatics* **65**, e57 (2019).
108. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
109. Kelley, D. R., Liu, B., Delcher, A. L., Pop, M. & Salzberg, S. L. Gene prediction with glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res.* **40**, e9 (2012).
110. Testa, A. C., Hane, J. K., Ellwood, S. R. & Oliver, R. P. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* **16**, 170 (2015).
111. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
112. Minos—a gene model consolidation pipeline for genome annotation projects. *GitHub* <https://github.com/EI-CoreBioinformatics/minos> (2019).
113. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
114. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
115. Kang, Y.-J. et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
116. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* **48**, 4.11.1–4.11.39 (2014).
117. Eilbeck, K., Moore, B., Holt, C. & Yandell, M. Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* **10**, 67 (2009).
118. Dainat, J. et al. AGAT: another gff analysis toolkit to handle annotations in any gtf/gff format. (Version v0.9.2). *Zenodo* <https://www.doi.org/10.5281/zenodo.6621429> (2022).
119. Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
120. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
121. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* <https://doi.org/10.12688/f1000research.7563.2> (2016).
122. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
123. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
124. Ritchie, M. E. et al. *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
125. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10**, 946–963 (2016).

126. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
127. Liu, R. et al. Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Res.* **43**, e97 (2015).
128. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2022).
129. Langfelder, P. & Horvath, S. Fast R functions for robust correlations and hierarchical clustering. *J. Stat. Softw.* **46**, 1–17 (2012).
130. Yu, G., Wang, L.-G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
131. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
132. Wijesooriya, K., Jadaan, S. A., Perera, K. L., Kaur, T. & Ziemann, M. Urgent need for consistent standards in functional enrichment analysis. *PLoS Comput. Biol.* **18**, e1009935 (2022).
133. Smyth, G. K., Michaud, J. & Scott, H. S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067–2075 (2005).
134. Amborella Genome Project. et al. The *Amborella* genome and the evolution of flowering plants. *Science* **342**, 1241089 (2013).
135. Lamesch, P. et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
136. Moreau, H. et al. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* **13**, R74 (2012).
137. Liu, S. et al. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat. Commun.* **5**, 3930 (2014).
138. Wang, X. et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* **43**, 1035–1039 (2011).
139. The International Brachypodium Initiative. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768 (2010).
140. Slotte, T. et al. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* **45**, 831–835 (2013).
141. Blanc, G. et al. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* **13**, R39 (2012).
142. Wan, T. et al. A genome for gnetophytes and early evolution of seed plants. *Nat. Plants* **4**, 82–89 (2018).
143. Bowman, J. L. et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304.e15 (2017).
144. Worden, A. Z. et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**, 268–272 (2009).
145. Ouyang, S. et al. The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* **35**, D883–D887 (2007).
146. Nystedt, B. et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).
147. The Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635–641 (2012).
148. Argout, X. et al. The genome of *Theobroma cacao*. *Nat. Genet.* **43**, 101–108 (2011).
149. Palenik, B. et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl Acad. Sci. USA* **104**, 7705–7710 (2007).
150. De Clerck, O. et al. Insights into the evolution of multicellularity from the sea lettuce genome. *Curr. Biol.* **28**, 2921–2933.e5 (2018).
151. Prochnik, S. E. et al. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329**, 223–226 (2010).
152. Wickell, D. et al. Underwater CAM photosynthesis elucidated by *Isoetes* genome. *Nat. Commun.* **12**, 6348 (2021).
153. Marchant, D. B. et al. Dynamic genome evolution in a model fern. *Nat. Plants* **8**, 1038–1051 (2022).
154. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
155. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
156. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
157. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
158. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
159. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
160. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611 (1965).
161. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1**, 50–60 (1947).
162. Müller, A. O., Blersch, K. F., Gippert, A. L. & Ischebeck, T. Tobacco pollen tubes—a fast and easy tool for studying lipid droplet association of plant proteins. *Plant J.* **89**, 1055–1064 (2017).
163. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
164. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
165. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
166. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
167. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).
168. Wang, Z. & Benning, C. *Arabidopsis thaliana* polar glycerolipid profiling by thin layer chromatography (TLC) coupled with gas-liquid chromatography (GLC). *J. Vis. Exp.* <https://doi.org/10.3791/2518> (2011).
169. Reich, M. et al. Fatty acid metabolism in the ectomycorrhizal fungus *Laccaria bicolor*. *New Phytol.* **182**, 950–964 (2009).
170. Miquel, M. & Browse, J. Arabidopsis mutants deficient in polyunsaturated fatty acid synthesis. Biochemical and genetic characterization of a plant oleoyl-phosphatidylcholine desaturase. *J. Biol. Chem.* **267**, 1502–1509 (1992).
171. Hornung, E. et al. Production of (10E,12Z)-conjugated linoleic acid in yeast and tobacco seeds. *Biochim. Biophys. Acta* **1738**, 105–114 (2005).

172. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
173. Clauw, P. et al. Leaf responses to mild drought stress in natural variants of *Arabidopsis*. *Plant Physiol.* **167**, 800–816 (2015).
174. Lu, Z. et al. Genome-wide DNA mutations in *Arabidopsis* plants after multigenerational exposure to high temperatures. *Genome Biol.* **22**, 160 (2021).
175. Suzuki, N. et al. ABA is required for plant acclimation to a combination of salt and heat stress. *PLoS ONE* **11**, e0147625 (2016).
176. Wang, L. et al. Differential physiological, transcriptomic and metabolomic responses of *Arabidopsis* leaves under prolonged warming and heat shock. *BMC Plant Biol.* **20**, 86 (2020).
177. Zhang, S.-S. et al. Tissue-specific transcriptomics reveals an important role of the unfolded protein response in maintaining fertility upon heat stress in *Arabidopsis*. *Plant Cell* **29**, 1007–1023 (2017).
178. Elzanati, O., Mouzeyar, S. & Roche, J. Dynamics of the transcriptome response to heat in the moss, *Physcomitrella patens*. *IJMS* **21**, 1512 (2020).
179. Jahan, A. et al. Archetypal roles of an abscisic acid receptor in drought and sugar responses in liverworts. *Plant Physiol.* **179**, 317–328 (2019).
180. Lagercrantz, U. et al. DE-ETIOLATED1 has a role in the circadian clock of the liverwort *Marchantia polymorpha*. *New Phytol.* **232**, 595–609 (2021).
181. Wu, T.-Y. et al. Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. *Nat. Plants* **7**, 787–799 (2021).
182. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
183. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
184. Almeida-Silva, F. & Venancio, T. M. cageminer: an R/Bioconductor package to prioritize candidate genes by integrating genome-wide association studies and gene coexpression networks. *In Silico Plants* **4**, diac018 (2022).
185. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

Acknowledgements

We thank R. Heise for excellent technical support. J.d.V. thanks the European Research Council for funding under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 852725; ERC-StG 'TerreStriAL'). J.d.V., U.H., I.F. and H.B. are grateful for support through the German Research Foundation (DFG), on the grant SHOAL (514060973; VR132/11-1) and within the framework of the Priority Programme 'MAdLand – Molecular Adaptation to Land: Plant Evolution to Change' (SPP 2237; 440231723 VR 132/4-1; BU 2301/6-1; HO 2793/5-1; FE 446/14-1), in which T.P.R. and M.H. are PhD students and A.D., J.M.R.F.-J and I.I. partake as associate members. A.D. is grateful for being supported through the International Max Planck Research School (IMPRS) for Genome Science. J.M.R.F.-J. and T.P.R. gratefully acknowledge support by the PhD programme 'Microbiology and Biochemistry' within the framework of the 'Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences' (GGNB) at the University of Goettingen. P.S. was supported by the GGNB in frame of the PRoTECT programme at the University of Goettingen. T.I. acknowledges funding from DFG (GRK 2172-PRoTECT). M.M. is supported by Singaporean Ministry of Education grant

T2EP30122-0001. P.S. is grateful for support from the Studienstiftung des Deutschen Volkes. This work was further supported by the DFG through the infrastructure grant INST 211/903-1 FUGG for the used confocal microscope as operated by the Imaging Network of the University of Münster (RI_00497). We thank C. Gatz and G. Kriete for giving us access to the ImagMAX/L PAM in the Department of Plant Molecular Biology and Physiology. We are grateful to T. Friedl for supporting us with access to the facilities of the Department of Experimental Phycology and SAG Culture Collection of Algae, including cultivation facilities, the Olympus BX-60 microscope and the absorbance microplate reader Epoch (BioTek Instruments).

Author contributions

J.d.V. and M.L. conceived the project. J.d.V. coordinated the project with M.M. M.L. provided the plant material. J.M.R.F.-J., T.D., S.S., M.L., and T.P.R. performed the experimental work. A.D. carried out the computational analysis. O.V., J.M.R.F.-J., P.S., T.I., D.K. and G.H.B. performed the proteomics. D.K., C.H. and I.F. performed the lipid profiling. H.B. investigated the cell division patterns. M.H. and U.H. investigated the photomorphogenesis patterns. A.D. and R.S. built the web resources. J.d.V., A.D. and J.M.R.F.-J. contributed to writing the manuscript. J.d.V. organized the manuscript. All authors commented, discussed and provided input on the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41477-023-01491-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-023-01491-0>.

Correspondence and requests for materials should be addressed to Jan de Vries.

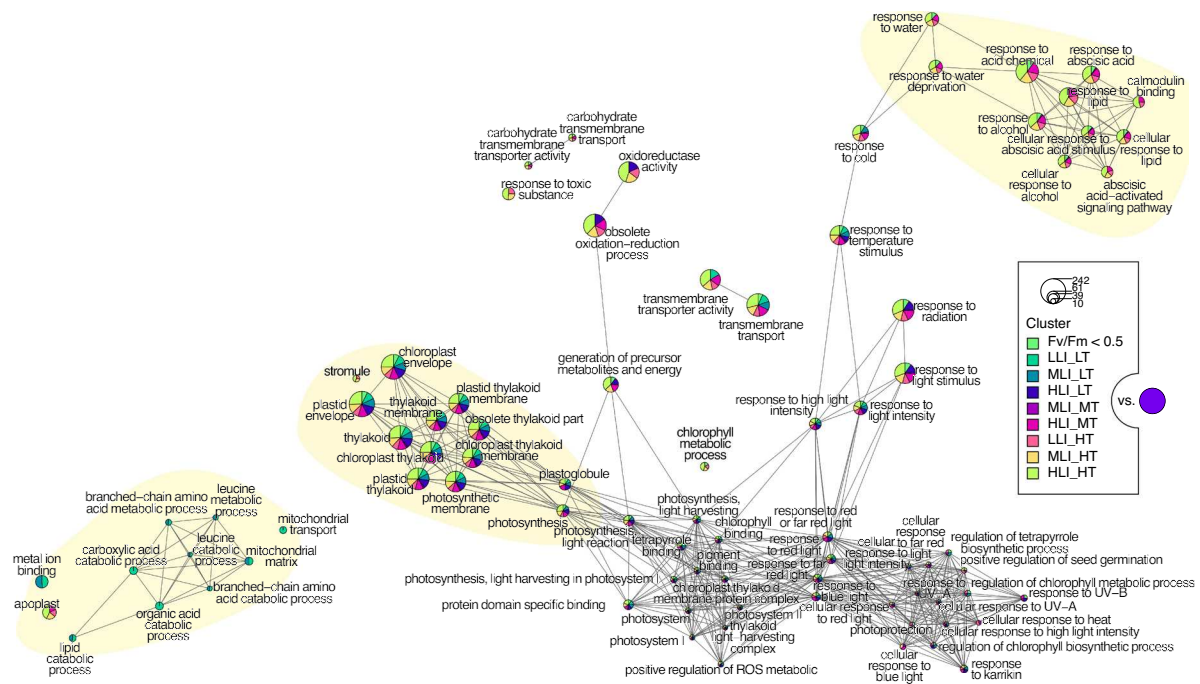
Peer review information *Nature Plants* thanks Xin Liu, Mingbing Zhou and Haim Treves for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

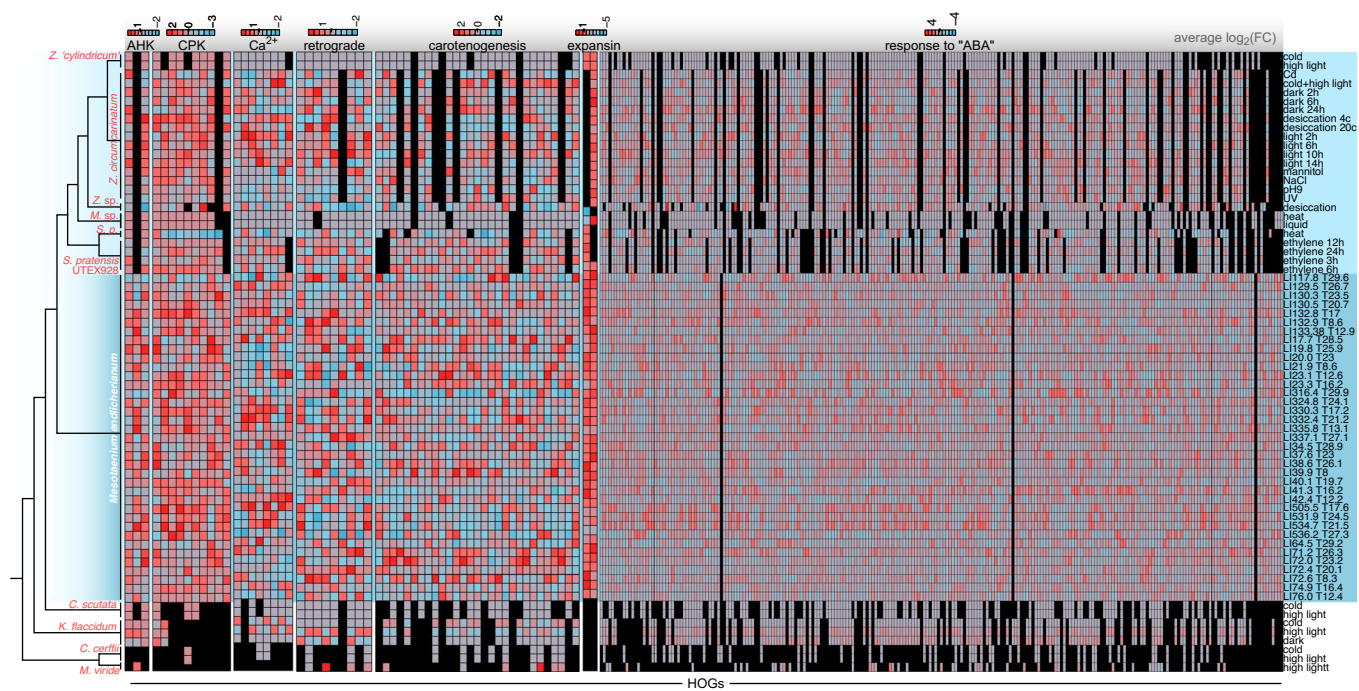
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

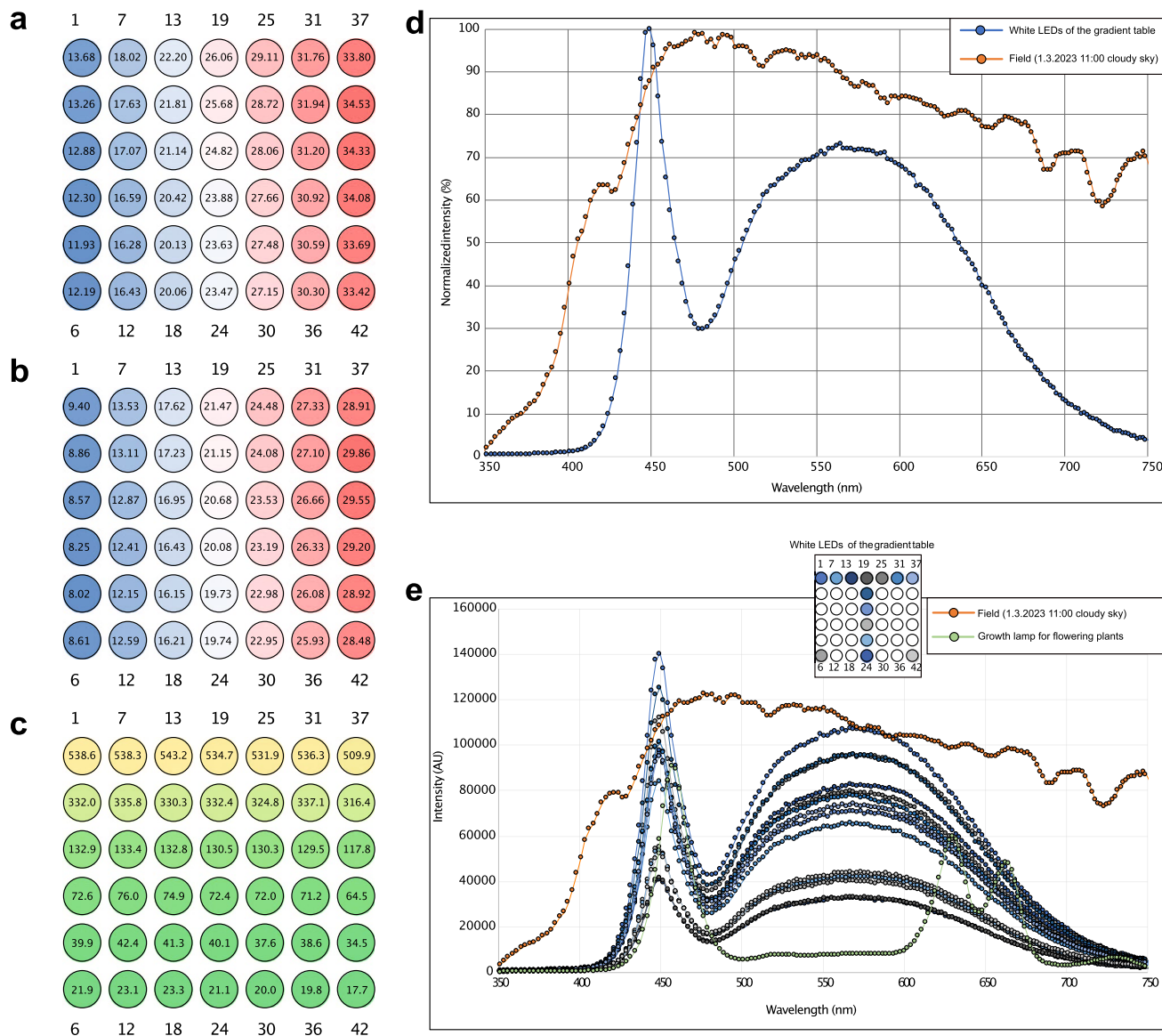


Extended Data Fig. 1 | Biological theme comparison summarizing all GO term enrichment analysis with adjusted p value ≤ 0.01 of DEGs against all genes that were expressed and passed the filtering in our analyses as background. The size of each circle is proportional to the count of each GO-term. Only the top 30 enriched terms are shown.



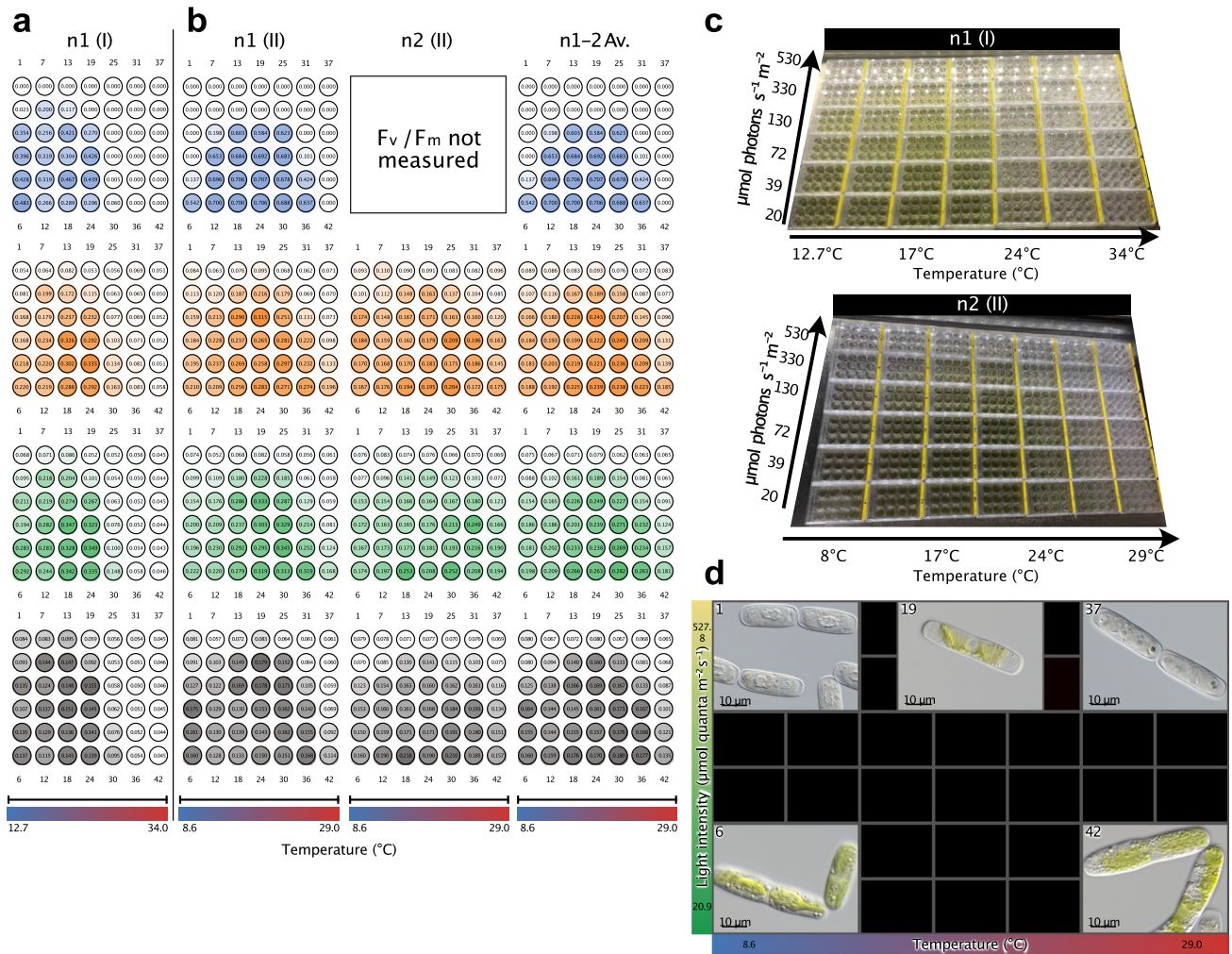
Extended Data Fig. 2 | Heat maps of average differential gene expression in $\log_2(\text{fold change})$ per HOG. From the strongest upregulation in red to the strongest downregulation in blue; black means that no HOG was found. The heat

maps were sorted by phylogeny (see the cladogram on the left) and treatment (written on the right); light teal highlights the data on Zygnematophyceae, dark teal on *Mesotaenium endlicherianum*.



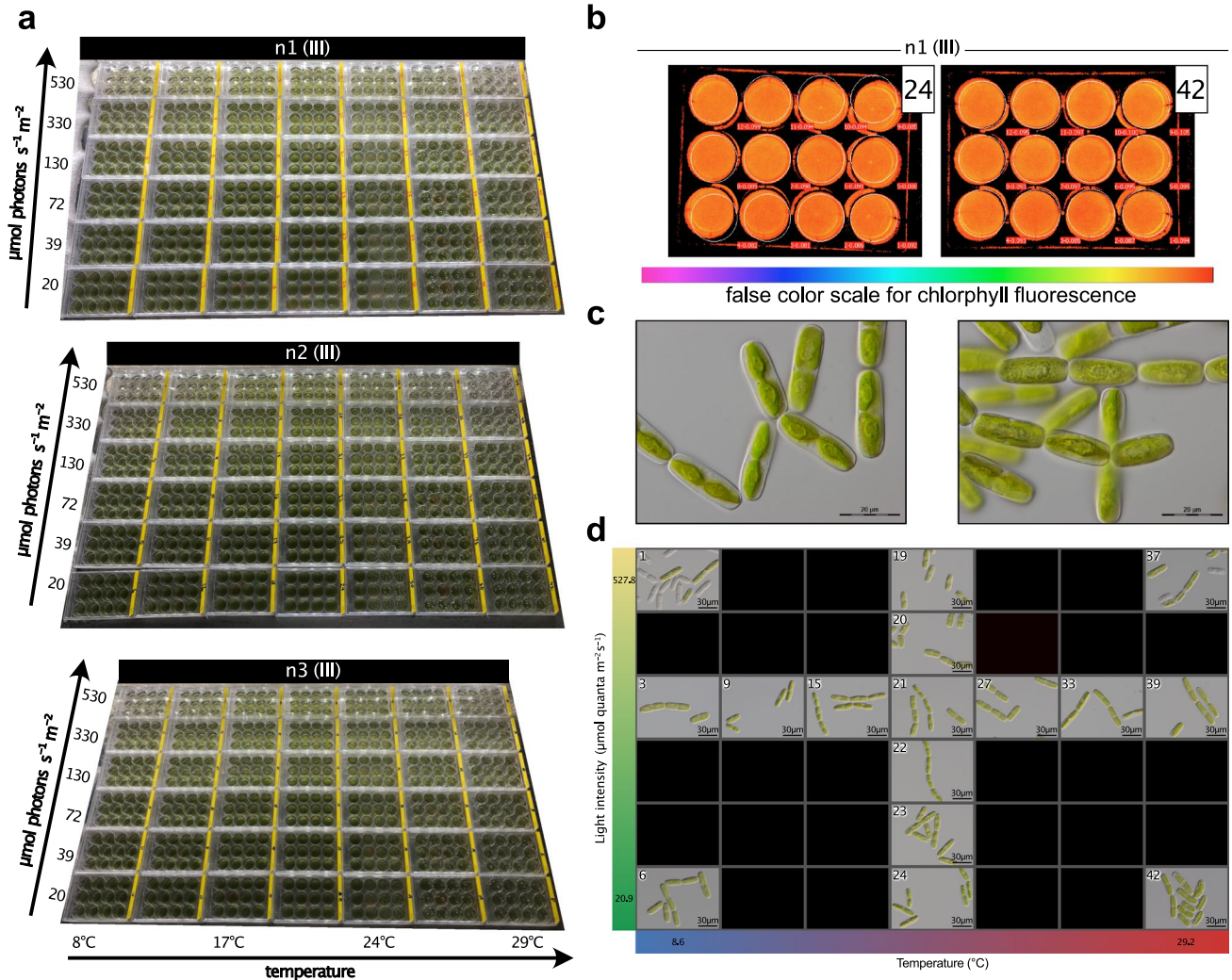
Extended Data Fig. 4 | Pre-experimental setups: temperature conditions comparison, light intensities, and light spectra. (a) Temperature conditions for the first experimental setup (n1(I)) depicted by a blue to red color gradient (see also supplementary table ST 1.4). (b) Temperature conditions for the second and the final experimental setup (n1(II), n2(III), n1(III), n2(III), and n3(III)) depicted by a blue to red color gradient (see also supplementary table ST 1.2). (c) Light intensity/irradiance values depicted by a green to yellow color gradient (see also

supplementary table ST 1.1). (d) Average light spectra of the gradient table (blue) assessed using SpectraPen (PSI, Brno, CZ) compared to a spectra assessed of natural sunlight (orange). (e) Light spectra of various plates of the gradient table (see overview) in various shades of blue assessed using SpectraPen (PSI, Brno, CZ) compared to a light spectrum assessed of natural sunlight (orange) and a light spectrum from a growth lamp used for flowering plants.



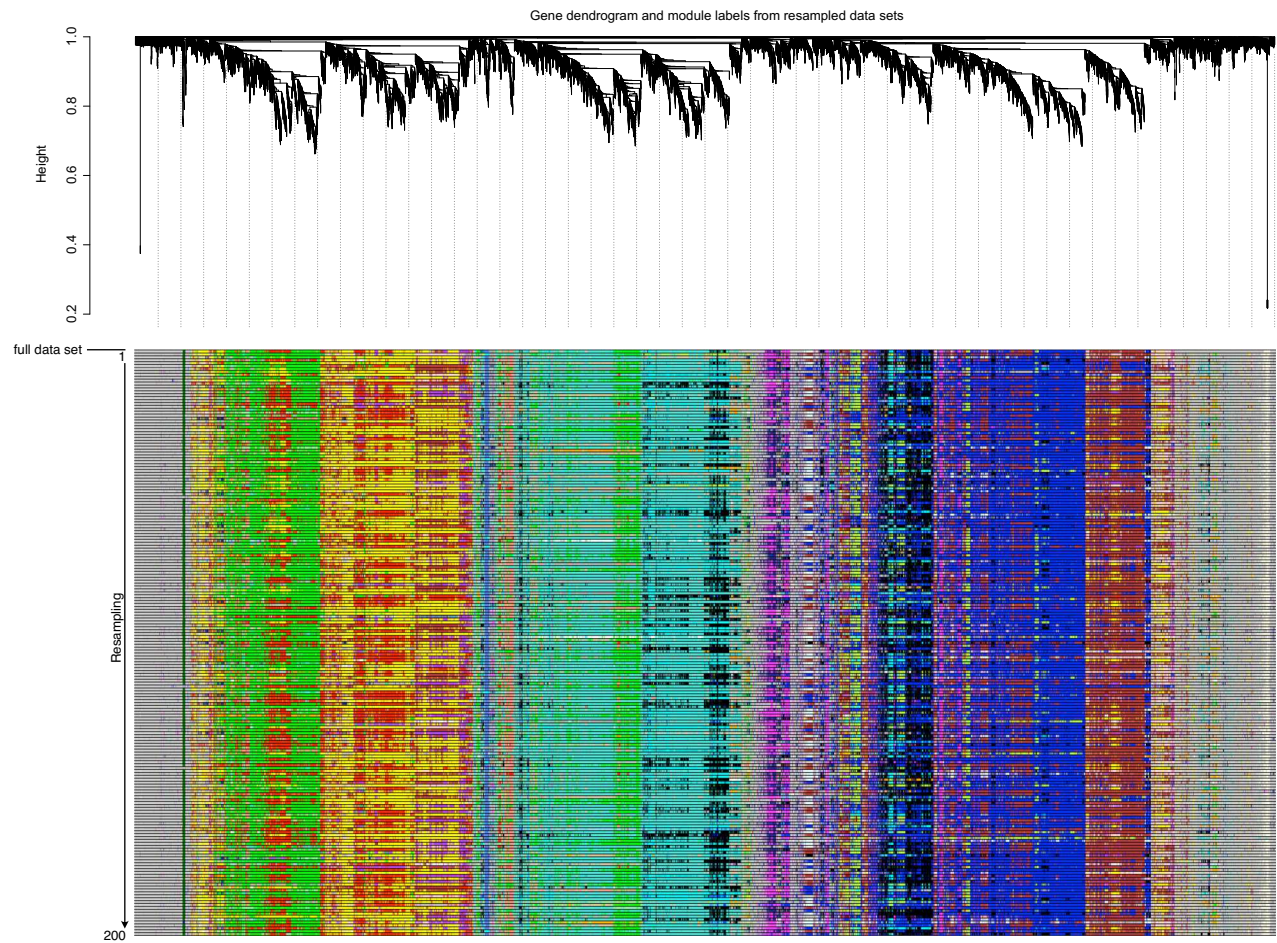
Extended Data Fig. 5 | Pre-experimental setup II: F_v/F_m, absorption, and morphology. (a) F_v/F_m values (blue gradient) and absorption values (orange, green, grey color gradient). Colors indicate measured wavelength: orange gradients = Absorption measured at λ 480 nm, green gradients = Absorption measured at λ 680 nm, grey gradients = Absorption measured at λ 750 nm) of the first pre-experimental setup (n1(I)) with temperature settings ranging from 12.7–34 °C. (b) F_v/F_m and absorption values of second pre-experimental setup (n1(II) and n2(II)) and averaged values (n1-2 Av.) with new temperature settings ranging from 8.6–29.0 °C. (c) Photographs of the pre-experimental setups n1(I) with

temperature conditions ranging from 12.7–34 °C and n2(II) with temperature conditions ranging from 8.6 to 29.0 °C after incubation on the table for 216 h (n1(I)) or 191 h (n2(II)) respectively. The photograph of pre-experiment n2(I) is not shown. (d) Differential interference contrast (DIC) micrographs of SAG 12.97 cells (pre-experimental setup n1(II)) under most extreme environmental conditions (four corners: samples 1, 6, 37, and 42) as well as under high irradiance 527.8 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ at 20.5 °C; for each well, at least 10 micrographs were taken, all showing similar phenotypes of the cells.



Extended Data Fig. 6 | Main-experimental setup (n1,2,3 (III)): Morphology and growth. (a) Photographs of the main experimental setups n1, n2, and n3 (III) with temperature conditions ranging from 8.6 to 29.0 °C after incubation on the table for 65 h. (b) Fm measurements (maximal fluorescence) using IMAGING-PAM in various table conditions, legend on the right is a false color gradient indicating fluorescence intensity. (c) Differential interference contrast (DIC) micrographs of SAG12.97 cells grown on C-Medium (growth conditions see methods: growth conditions prior to exposure to environmental conditions);

at least 10 micrographs were taken, all showing similar phenotypes of the cells. (d) Differential interference contrast (DIC) micrographs of SAG12.97 under most extreme environmental conditions (four corners: samples 1, 6, 37, and 42) as well as along an irradiance gradient at 21 °C (samples 19–24) and a temperature gradient at 130 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ (samples 3, 9, 15, 21, 27, 33, and 39); for each well, at least 10 micrographs were taken, all showing similar phenotypes of the cells.



Extended Data Fig. 8 | Module stability analysis. We performed 200 simulations by sampling from our input expression profiles under different conditions using the WGCNA package. Each row in the heat map represents a simulation and stable modules form a 'column' of similar color on the heat map.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Absorption were measured and averaged using software Gen5 v(2.0). Photophysiological measurements via the IMAGING PAM have been done via ImagingWinGigE (V2.32) software. Sequencing data was collected on an Illumina NovaSeq6000 platform operated by Novogene UK. wget (GNU Wget 1.14) was used to download genomic sequences from Phytozome, Novogene, DDBJ, Figshare, Fernbase, and TAIR databases. LC-MS method programming and data acquisition for raw proteome data was performed with the XCalibur 4.0 software (Thermo Fisher Scientific).

Data analysis

We installed all packages (and their dependencies) using conda if possible. Otherwise, we followed the installation guide offered by the authors of the tool. For R packages, we installed them from, in this order of priorities, from Cran repository, BioCmanager, or the tool GitHub instructions. All codes used to perform computational analyses are available in our GitHub repository: https://github.com/deVries-lab/Response_to_a_gradient_of_environmental_cues_in_mesotaenium_endlicherianum

Absorption were measured and averaged using software Gen5 v(2.0). Photophysiological measurements via the IMAGING PAM have been done via ImagingWinGigE (V2.32) software.

For statistical analyses of absorption and Fv/Fm values and temperature/light cluster analysis, we used these set of tools: Kruskal-Wallis test with post hoc test Fisher's least significant difference using R (4.1.3). P-values were Bonferroni corrected and grouped into significant groups using R packages 'agricolae' (v1.3-5) and 'dplyr' version (v1.0.9), pheatmap (v1.0.12), from the factoextra package (v1.0.7) we used eclust function with clustering function 'kmeans', with number of clusters set to six and for hierarchical clustering 'euclidean' was used as distance measure.

The quality of raw RNA-Seq reads were checked with FastQC (v0.11.9) and summarized with MultiQC (v1.11). Reads were trimmed and

filtered via Trimmomatic (v0.36) with these parameters:

```
("ILLUMINACLIP: novogene_adapter_sequences_Trimmomatic.fa:2:30:10:2:True
418 LEADING:26 TRAILING:26 SLIDINGWINDOW:4:20 MINLEN:36")
```

The quality of trimmed and filtered reads were again checked by FastQC (v0.11.9) and MultiQC (v1.11).

For genome re-annotation we used a diverse set of bioinformatic tools including:

HISAT2 (v2.2.1), StringTie(v2.1.5), Busco (v5.3.2), REAT (v0.6.1, <https://reat.readthedocs.io/en/latest/>), scallop (v0.10.5), Portcullis (v1.2.4), Mikado (v2.3.4), SPALN (v2.4.7), Augustus (v3.4.0), SNAP (v2006-07-28), Glimmer (v0.3.2), CodingQuarry (v2.0), EvidenceModeler (v1.1.1), Minos (v1.8.0; <https://github.com/EI-CoreBioinformatics/minos>), Diamond (v0.9.34), Kallisto (v0.46.2), CPC2 (v0.1), maker (v3.01.04), interproscan (v0.9.2), agat (v0.9.2), eggNOGmapper (v2.1.8)

For RNA-Seq quantification and differential gene expression analyses we used this set of software:

Snakemake((7.7.0), Kallisto (v0.45.0), R (v4.2.0), tximport (v1.24.0) with "lengthScaledTPM" option, tidyverse (v1.3.1), edgeR (v3.38.1) "calcNormFactors method=TMM", limma (v3.52.2), ggplot2 (v3.3.6), pheatmap (v1.0.12) with "clustering_distance_rows = "euclidean" and clustering_distance_cols = "euclidean"", clusterProfiler (v4.4.4) with p-value and q-value cutoff of 0.01 and all genes that passed low-expression filtering as background universe

We performed Weighted gene co-expression analyses via:

WGCNA (v1.71), tidyverse (v1.3.1), clusterProfiler (v4.4.4) with p-value and q-value cutoff of 0.05 and all genes that passed low-expression filtering as background universe, eggNOGmapper (v2.1.8), BLAST (v2.11.0+)

We performed Phylogenetic analyses using:

BLAST (v2.11.0+), MAFFT (v7.490), IQ-TREE (v1.5.5), ModelFinder (integrated in IQ-TREE multicore version 1.5.5 for Linux 64-bit built Jun 2 2017) according to Bayesian Information Criterion and 1000 ultrafast bootstrap replicates; 1000 ultrafast bootstrap replicates were carried out 100 Felsenstein bootstraps. We colored phylogeny trees via ggtree (v3.9.0)

Protein domains were predicted using Interproscan (v5.59-91.0)

For transcriptome analyses of published data from other algae, we used Trinity (v2.15.1), Transdecoder (v5.7.0), and ComBat-seq. For orthogroup analyses we used BioNERO and Orthofinder

Proteome data were analyzed using Max Quant software version 1.6.2.10 (Cox and Mann, 2008) and further processed using Perseus (1.6.2.2) software.

Gas-chromatograms were analysed using the ChemStation software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All RNAseq reads have been uploaded to NCBI SRA and can be accessed under Bioproject PRJNA832564 and SRA accessions SRR18936040 to SRR18936170.

Furthermore, data can be interactively explored at <https://mesotaenium.uni-goettingen.de>

Proteomic data have been uploaded to EMBL-EBI PRIDE (accession PXD037847).

On Zenodo, we have deposited: (i) raw light and confocal micrographs generated, e.g. for lipid droplet assessment in Mesotaenium and pollen tubes <https://doi.org/10.5281/zenodo.7921367> and (ii) Raw and visualized phylogenetic data <https://doi.org/10.5281/zenodo.7950653>

The additional previously published RNAseq datasets that were used for comparisons are: (i) *A. thaliana*: SRR2302908 to SRR2302919, ERR754084, ERR754066, ERR754077, ERR754069, ERR754087, ERR754064, ERR754059, SRR7659142, SRR7659143, SRR7659144, SRR7659145 to SRR7659150, SRR5197904, to SRR5197909; (ii) *M. polymorpha*: SRR12076853, SRR12076855, SRR12076857, SRR12076859, SRR12076861, SRR12076863, SRR12076865, SRR12076867, SRR12076869, SRR12076871, SRR12076873, SRR12076875, SRR12076877, SRR12076879, SRR12076917 to SRR12076925, SRR15186078 to SRR15186125, DRR093991 to DRR093996; (iii) *P. patens*: SRR1824306 to SRR1824320, SRR10235460 to SRR10235483, SRR787291, SRR787292, SRR787293, SRR787294, SRR787295; (iv) *Z. circumcarinatum* SAG698-1b: SRR24939299, SRR24940177, SRR24909175, SRR24757807, SRR24757829, SRR24757830, SRR24757831, SRR24205691 to SRR24205702, SRR24286545 to SRR24286562, SRR24576622, SRR24576623, SRR24385702, SRR24450996, SRR24450997, SRR24451196, SRR24480449, SRR24707416, SRR24707417, SRR24952091, SRR21891679 to SRR21891705; (v) *C. cerffii* (at the time, *C. atmophyticus*, see ref. 97): SRR5949009, SRR5949013 to SRR5949016, SRR5949027 to SRR5949030; (vi) *C. scutata*: SRR5948993, SRR5948995 to SRR5948998, SRR5949001, SRR5949004, SRR5949005, SRR5949007; (vii) *K. flaccidum*: SRR5949010, SRR5949011, SRR5949012, SRR5990072 to SRR5990080; (viii) *M. viride*: SRR5949021 to SRR5949026; (ix) *Mougeotia* sp. MZCH240: SRR9083681, SRR9083682, SRR9083688, SRR9083692 to SRR9083701; (x) *S. pratensis* MZCH10213: SRR9083685, SRR9083686, SRR9083687, SRR9083689, SRR9083690, SRR9083696; (xi) *S. pratensis* UTEX928: SRR4018077 to SRR4018100; (xii) *Z. circumcarinatum* SAG698-1a: SRR5948999, SRR5949000, SRR5949002, SRR5949003, SRR5949006, SRR5949008, SRR5949017, SRR5949018; and (xiii) *Z. circumcarinatum* SAG2419: SRR6047298, SRR6047299, SRR6047302 to SRR6047305.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Each analysis involved millions of pooled cells (cultures were set to a density of 20,300,000 cells/ml), all can be assumed to behave similarly (as they were vegetative cells from the same starting culture). For each experiment, 504 Mesotaenium endlicherianum samples were analyzed, pooled to 42 samples for RNAseq and averaged to 42 physiological/growth data points (measurements for all 504 samples are provided in the supplement). Sequencing was then performed to a depth that was chosen based on approaching saturation level (based on obtaining differential expression patterns among the given number of genes).
Data exclusions	No data were excluded.
Replication	For the physiological and morphological analysis as well as RNAseq, three independent experiments were conducted. Preexperiments for 8 days were carried out two times, 31 March 2021 to 8 April 2021 and 26 May 2021 to 4 June 2021. For the final 3 day-lasting experiments for RNAseq combined with physiological assessment the gradient table was run three times independently in November 2021 (8 Nov. 21 to 11. Nov. 21; 15. Nov. 21 to 18. Nov. 21; 23. Nov. 21 to 26. Nov. 21). In this final setup, all attempts at replication were successful. For proteomic, analysis of two samples with a successfully enriched lipid droplet fraction and the corresponding total extract fraction, both isolated from the alga Mesotaenium endlicherianum, were used. For lipid droplet counts, hundreds of cells were assessed, indicated in the figure.
Randomization	All experiments are based on a random selection of millions of cells from a liquid culture. The start culture was one homogenous culture that was equally and distributed. A random selection of millions of cells thus ended up in one plate that was exposed to a certain condition. Prior to start, all plates were thus equal.
Blinding	Blinding was not relevant for this study. It is irrelevant for the bioinformatics because we worked with all versus all comparisons, unsupervised methods and all pipelines are fully transparent on GitHub. All cell-based evaluation is quantifiable and unambiguous (counts of lipid droplets). Further, the information is fully provided and re-evaluable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Mesotaenium endlicherianum (Zygnematophyceae, Streptophyta): The alga has the collection number SAG 12.97. The biomaterial provider is the Experimental Phycology and Culture Collection of Algae in Göttingen, Germany. The alga was originally isolated in Portugal, Quiaios, Lagoa das Bracas, plankton, Lat./Long.(Precision): 40.243191 / -8.80488.
Authentication	Authentication was carried out directly by the biomaterial provider, the Experimental Phycology and Culture Collection of Algae in Göttingen, Germany, via microscopy and genetic markers.
Mycoplasma contamination	n/a
Commonly misidentified lines (See ICLAC register)	n/a

4 Chapter II: Rebuilding Streptophyte Evolution with Comparative Genomics

4.1 **Publication IV:** Crossroads in the evolution of plant specialized metabolism

This review paper was published online in the Journal “Seminars in Cell and Developmental Biology” in March 2022. The full article as well as supplementary datasets can be found under also be found online:

<https://doi.org/10.1016/j.semcdb.2022.03.004>

Contribution of Janine Fürst-Jansen, third author

J. M. R. Fürst-Jansen designed and organized figure 1 and contributed to the content of figures 2 and 3. She wrote a paragraph of the chapter “Enzyme promiscuity, shreds of polymers, and the biosynthesis of phenylpropanoid-derived compounds” and critically read and revised the manuscript.



Contents lists available at ScienceDirect

Seminars in Cell and Developmental Biology

journal homepage: www.elsevier.com/locate/semcdb

Crossroads in the evolution of plant specialized metabolism

Tim P. Rieseberg^{a,1}, Armin Dadras^{a,2,3}, Janine M.R. Fürst-Jansen^{a,3,4},
 Amra Dhabalia Ashok^{a,5}, Tatyana Darienko^{a,6}, Sophie de Vries^{a,3,7}, Iker Irisarri^{a,b,3,8},
 Jan de Vries^{a,b,c,*}⁹

^a University of Goettingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany

^b University of Goettingen, Campus Institute Data Science (CIDAS), Goldschmidtstr. 1, 37077 Goettingen, Germany

^c University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany

ARTICLE INFO

Keywords:

Charophytes
 Earliest land plants
 Specialized metabolism
 Apocarotenoids
 Phenylpropanoids
 Exaptation
 Plant evolution
 Plant terrestrialization
 Streptophyte algae
 Stress physiology

ABSTRACT

The monophyletic group of embryophytes (land plants) stands out among photosynthetic eukaryotes: they are the sole constituents of the macroscopic flora on land. In their entirety, embryophytes account for the majority of the biomass on land and constitute an astounding biodiversity. What allowed for the massive radiation of this particular lineage? One of the defining features of all land plants is the production of an array of specialized metabolites. The compounds that the specialized metabolic pathways of embryophytes produce have diverse functions, ranging from superabundant structural polymers and compounds that ward off abiotic and biotic challenges, to signaling molecules whose abundance is measured at the nanomolar scale. These specialized metabolites govern the growth, development, and physiology of land plants—including their response to the environment. Hence, specialized metabolites define the biology of land plants as we know it. And they were likely a foundation for their success. It is thus intriguing to find that the closest algal relatives of land plants, freshwater organisms from the grade of streptophyte algae, possess homologs for key enzymes of specialized metabolic pathways known from land plants. Indeed, some studies suggest that signature metabolites emerging from these pathways can be found in streptophyte algae. Here we synthesize the current understanding of which routes of the specialized metabolism of embryophytes can be traced to a time before plants had conquered land.

1. Streptophyte terrestrialization and the challenge of living on land

The surface of our planet is covered by a monophyletic group of organisms whose cumulative biomass dwarves that of all other life: the land plants (Embryophyta; [1]). Embryophytes, to which we will henceforth simply refer to as land plants, are not the only photosynthetic eukaryotes on land. Algae from various lineages dwell on land, including

chlorophytes, diatoms and many more [2–4]. The terrestrial macroflora is however solely composed of land plants.

Land plants emerged from within the clade of Streptophyta about a little more than 500 million years ago [5]; the clade Streptophyta consists of land plants and streptophyte algae (see Box 1 for details). It thus was an ancestral freshwater and terrestrial streptophyte alga that gained a foothold on land and brought about the series of fateful events that are coined as ‘plant terrestrialization’ [6,7]. During this process, the earliest

* Corresponding author at: University of Goettingen, Institute for Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany.

E-mail address: devries.jan@uni-goettingen.de (J. de Vries).

¹ 0000-0003-3548-8475.

² 0000-0001-7649-2388.

³ contributed equally.

⁴ 0000-0002-5269-8725.

⁵ 0000-0001-5787-6941.

⁶ 0000-0002-1957-0076.

⁷ 0000-0002-5267-8935.

⁸ 0000-0002-3628-1137.

⁹ 0000-0003-3507-5195.

<https://doi.org/10.1016/j.semcdb.2022.03.004>

Received 14 October 2021; Received in revised form 17 February 2022; Accepted 4 March 2022

Available online 13 March 2022

1084-9521/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Box 1

Streptophyte algal diversity.

Streptophyta forms a monophyletic group comprising land plants (embryophytes) and the streptophyte algae (short, streptophyte algae). Currently, streptophyte algae represent two major grades: (1) KCM containing the classes Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae and (2) ZCC the classes Zygnematophyceae, Coleochaetophyceae, and Charophyceae [50,266–268]. The ZCC grade and land plants form the monophylum Phragmoplastophyta (“Core”-Streptophyta) [6,50,115,269,270]. All recent phylogenomic analyses agree on that the Zygnematophyceae are the algal sister lineage to land plants [50,268–270]. The current status about the streptophyte algae is summarized in the following table (numbers were extracted from <http://www.algaebase.org>).

		# genera	# species	Morphotype	Ecology
KCM	Mesostigmatophyceae	1	2	monadoid	freshwater
	Chlorokybophyceae	1	5	sarcinoid	terrestrial
	Klebsormidiophyceae	5	41	trichal + sarcinoid	terrestrial/freshwater
ZCC	Charophyceae	3	> 110	trichal	freshwater/brackish
	Coleochaetophyceae	2	29	trichal + parenchymatous	freshwater
	Zygnematophyceae	45	> 4200	trichal + coccoid	terrestrial/freshwater

The classes Mesostigmatophyceae and Chlorokybophyceae form sister groups to each other [266] and are most distantly related to the other algal classes of the Streptophyta. Representatives of both classes are rare organisms and only recognized by ultrastructure of the flagellated cells as members of the Streptophyta [271,272]. In both classes, sexual reproduction is unknown (for an overview, see also [273]). Curiously, some *Spirotaenia*, originally assigned as member of Zygnematophyceae based on morphological features such as special type of sexual reproduction (conjugation) and absence of any flagellated stages, appear closely related to Chlorokybophyceae according to SSU and *rbcL* phylogenies [274] and phylogenomics [275]. Recently, *Chlorokybus* was found to entail not only one species but a cryptic species complex, encompassing at least five extant members [275].

The class Klebsormidiophyceae are filamentous and single-celled algae that represents the sister group to the Phragmoplastophyta [269, 276–279]. The genus *Klebsormidium* is widely distributed in almost all non-marine habitats and is able to form biological soil crusts. Despite the controversial taxonomy within this genus, *Klebsormidium* became a well-researched organism. Other genera of Klebsormidiophyceae are very rare algae (except for *Interfilum*).

The Charophyceae (stoneworts) include freshwater (occasionally brackish) algae with complex macroscopic thalli composed of a main axis with whorled branches characterized by growth of apical meristematic cells. The sexual reproduction is a specialized oogamous mode with oogonia and antheridia surrounded by sterile cells. Charophyceae are well represented in the fossil record with a large diversity extending back to the Silurian [273,280,281]. According to McCourt et al. [280], currently six extant genera in the family Characeae included in the order Charales are recognized. Two additional orders and a large number of genera and families are only known from the fossil records [282,283]. Two genera (*Chara*, *Nitella*) contain more than 100 described species each, with a third (*Tolypella*) containing several dozen taxa. The family Characeae contains all living charophytes with generally a worldwide distribution. Some taxa have very restricted distribution and are even endemic. In general, dioecious taxa are narrowly distributed or endemic, whereas monoecious taxa are usually widely distributed [284].

The class Coleochaetophyceae includes two genera *Coleochaete* and *Chaetosphaeridium*. The genus *Coleochaete* comprises algae with branched filaments or sometimes form complex discoid parenchymatous thalli [285]. Some thallus cells bear distinctive sheathed hairs [286]. Some species of *Coleochaete* form corticated zygotes, which often retain in the mother organism. The cytokinesis occurs by phragmoplast formation and presence of plasmodesmata similar to the land plants [287–289]. The zygote possesses sporopollenin, a highly resistant substance found in the outer wall of pollen [43]. Only eight *Coleochaete* species are available in culture and deposited in public culture collections [290]. *Chaetosphaeridium* is the second genus placed in the Coleochaetophyceae, based on molecular phylogenetic studies [276,277], morphological and cytological features, such as the presence of typical sheathed hairs and similar chloroplast structure [281,290,291]. This genus consists many loosely connected globose or flask-shaped cells imbedded in a gelatinous matrix. Vegetative cells bearing long sheathed hairs. Seven species of *Chaetosphaeridium* are described, but only two *C. pringsheimii* (type species of genus) and *C. globosum* are isolated and deposited in the public culture collections.

The Zygnematophyceae, also known as conjugating green algae, is the most species-rich and morphologically diverse lineage of the Streptophyta [281]. Morphologically, representatives of this class are very diverse, ranging from unicellular coccoid and small colonial forms (Desmidiaceae) to filaments (Zygnematales such as *Spirogyra*, *Zygnema* etc). The sexual reproduction occurs by a unique process of conjugation, involving fusion of non-motile gametes. The absence of flagellated reproductive stages, plasmodesmata and basal bodies also sets this class apart from the other streptophyte algal classes. Traditionally, the class was divided into the Zygnematales and the Desmidiaceae based primarily on differences in cell wall structure, but molecular phylogenies have shown that the Zygnematales (characterized by smooth non-ornamented cell walls) is a paraphyletic assemblage that gave rise to the monophyletic Desmidiaceae [281,292–295].

Zygnematophyceae occur in a wide variety of freshwater habitats, such as ephemeral pools, ponds, lakes, marshes, bogs, artificial habitats on every continent around the world. Within a given habitat, species often show preference for microhabitats. Most conjugating green algae are benthic or periphytic and grow on surfaces or occasionally attached to substrates by means of rhizoids or mucilage [296,297]. Some Zygnematales could be found in extreme habitats such on snow and ice (*Ancylonema*, *Cylindrocystis*, and *Mesotaenium*; [298]). *Cylindrocystis* has also been found in desert crust communities [3,299]. The filamentous *Zygogonium* is found in very acid pools and rivers (pH < 3) (e.g., [300]) or in High Alpine soils [301]. In general, desmids prefer slightly acidic waters (pH 4–7), such as pools in acid peat bogs [297,302]. The Desmidiaceae are represented by four families, the Closteriaceae, Gonatozygaceae, Peniaceae, and Desmidiaceae, which is the largest family [296]. The order Zygnematales traditionally included the families Zygnemataceae and Mesotaeniaceae [303]. The most prominent genus of this family is *Spirogyra* with around 300 morphospecies. Molecular phylogenetic studies indicate that the families belonging to the Zygnematales are not monophyletic and morphological features are often not correlated with phylogeny [292,293,304].

land plants (i.e. the first representatives of the monophylum Embryophyta; see also discussions in [8]) acquired a set of shared traits that distinguish embryophytes from all other streptophytes. One of the key aims of studying the early evolution of land plants is defining what these traits were and when they came about. Understanding the evolutionary origin of the defining traits of land plants holds the promise that we attain an understanding of the singularity of plant terrestrialization.

It is human to ask for reason and meaning. In the context of this manuscript, a looming question is “Why did land plants emerge only once?”. In (evolutionary) biology the answer is however always the same: because it worked. Or rather because it did *not* ‘not work’. When we apply this reasoning to the question of plant terrestrialization one approach is to ponder the challenges that were faced during plant terrestrialization. And challenges are aplenty in the terrestrial habitat.

Terrestrial challenges can be divided into two major classes—biotic and abiotic. Strategies to face biotic stressors are found in the whole green lineage and the conquest of land brought the need to adapt to new terrestrial pathogens [9]. This issue has been discussed and reviewed several times in recent years [9–12]. When comparing the physics of different habitats of the green lineage the conditions are considerably harsher in a terrestrial habitat than in a marine habitat with (shallow) freshwater habitats as an intermediate: For example, the fluctuations in environmental temperatures hinge on three major components: thermal conductance, capacity of the medium, and the size of the system. Temperature changes pronouncedly faster in air than in water; further, the size of a freshwater habitat is likely to be much smaller than that of a marine habitat. The same ranking holds true for the availability of water and osmotic pressure. The overall concentration of small molecules is rather constant in a given marine environment. In freshwater habitats the amount of water might vary quite drastically throughout the year and consecutively the osmotic pressure. For terrestrial plants this challenge is even more profound. Another important factor is irradiance. The quantity and quality of light differ significantly between all habitats. This is influenced mainly by the refractive index of the medium and the distance the light needs to travel through. Hence, the quantity and the spectral width of light is decreasing with increasing water depth (approximately 10 m below the water surface the red light is completely gone [13]).

All organisms need to respond to abiotic and biotic challenges. These include, for example, the modification of membranes under temperature stress and the production of phytoalexins under pathogen attack. Not least due to their sessile nature, land plants have a unique suite of molecular traits for responding to their environment (e.g., for unique means of modifying membranes with relevance to temperature stress, see [14]). And many of these traits hinge on their specialized metabolism. Streptophyte algae, which are phylogenetically the closest algal relatives of land plants, occupy a diversity of habitats: they occur in freshwater, and can grow hydroterrestrial as well as (aero-)terrestrial ([15]; see Box 1). As such, those lineages of streptophyte algae that have a foothold in the terrestrial environment obviously experience the same stressors as land plants at least during some time in their lives (see discussions in [8]). In combination with their close phylogenetic relationship, it is reasonable to argue that streptophyte algae and land plants will share a core set of routes of their specialized metabolisms to face similar biotic and abiotic terrestrial stressors.

Herein, we focus on the mitigation and sensing of abiotic stressors and dissect which routes of the specialized metabolism of embryophytes are already present in the closest algal relatives of land plants—the streptophyte algae—allowing for an inference of the routes that were present in their shared common ancestor(s).

2. Protective compounds and the diversification of plant specialized metabolism on land

Specialized metabolism is responsible for a wide variety of molecules that occur only in some organismal groups where they might or might

not be essential. Pathways of the specialized metabolism share several enzymes and compounds with primary metabolism, which calls for either (i) an evolutionary origin by branching from primary metabolism or (ii) for a shared evolutionary origin of both metabolisms. For example, phenylalanine, a necessary building block of proteins, is the first compound of the phenylpropanoid pathway; the primary metabolites isopentenyl pyrophosphate and dimethylallyl pyrophosphate are the source of isoprenoids; and caffeine and purine alkaloids derive from purine nucleotides that make up the DNA [16]; see also [17]. The manifold ties between primary and specialized metabolism are hardly surprising, given the interconnectivity of most primary metabolic pathways (e.g., the biosynthesis of tryptophan and histidine are interconnected to nucleotide biosynthesis and nitrogen assimilation pathways; see also [18]) and the multiple functions that primary metabolites often have in an organism (e.g., amino acids have many functions besides being protein monomers; [19]). Moreover, specialized metabolism is not merely a “dead end” branching off from primary metabolism, but often connects back to primary metabolism. An example is the retrograde flow of sulfur from glucosinolates to cysteine in *Arabidopsis* [20].

Under certain conditions, specialized metabolites provide selective advantages. Thus, it is unsurprising that in contrast to the (near) universality of core metabolism, plant specialized metabolism is highly diverse (between 200,000–1 million specialized metabolites exist compared to ~1000 primary metabolites) and shows a patchy distribution across the Tree of Life [17,18]. In plants and algae, specialized metabolites have crucial functions ranging from regulatory signals (e.g., phytohormones) to structural polymers (e.g., lignin) involved in any biological process including the protection against biotic and abiotic stresses. Pathways of specialized metabolism consist of multiple reactions carried out by a diverse range of enzymes from different gene families. As such they can be considered complex traits that offer a large working surface to be shaped by evolutionary forces and as such can hold an adaptive advantage to a new or challenging environment [17]. One of the most drastic suites of such adaptations to a new environment must have occurred during the conquest of land by plants due to the stark differences in habitats as illustrated above.

Indeed, several processes seen in extant land plants must have occurred to accomplish the water-to-land transition. An adequate acclimation and stress response are not the *only* requirements for succeeding on land. Yet, they are a necessity (see also [8]). The challenges that the algal lineage that conquered land had to face—and that all extant land plants still experience on a daily basis—include increased irradiance, temperature shifts, as well as recurring drought and flooding [7]. They also include the need to identify, communicate, and respond to other organisms [8–11,21–27]. The biosynthesis of protective compounds, such as phenylpropanoids and their derivatives, to minimize the damage caused by, for example, excessive light and the resulting reactive oxygen species (ROS), was a cornerstone in the evolution of land plants [8,18,28,29]. In the following sections we will start with the commonalities and differences in the specialized metabolism of streptophyte algae and embryophytes to dissect the evolutionary routes of plant specialized metabolism and its possible role in the conquest of land. Afterwards, we discuss the evolutionary forces that act upon specialized metabolisms to dissect how the picture of extant specialized metabolism came about.

3. Enzyme promiscuity, shreds of polymers, and the biosynthesis of phenylpropanoid-derived compounds

Some enzymes possess more than one catalytic activity and participate in various pathways, such as detoxification enzymes or aminotransferases [30]. These are known as multifunctional enzymes. Enzymes are also intrinsically promiscuous, meaning that they normally catalyze non-canonical reactions at low levels; the so-called underground metabolism [31]. Various mechanisms are responsible for this promiscuity, including substrate permissiveness, mechanistic elasticity,

and concomitant product diversity [18]. The “metabolic noise” naturally occurring at basal levels in the cell is the raw metabolic material onto which natural selection can act, selecting certain reactions and metabolites given the right conditions. It is important to note that the selection of metabolites and reactions always occurs at given cellular contexts (i. e., presence of some substrates and products at given concentrations), and therefore an enzyme’s substrate specificity is not universal but context-dependent. Intrinsic changes such as cell-specific expression, subcellular localization of metabolites, and extrinsic environmental changes will affect the cellular context—and therefore the kinetic optima for enzyme specificities. In fact, environmental variations can lead to variation in an organism’s fitness landscape [32]. The flexibility of metabolic pathways allows populations to more freely maneuver fitness landscapes and as such makes such pathways the target of adaptive evolution. Multifunctional enzymes and especially enzymatic promiscuity are hence key aspects for the evolvability of metabolic pathways; and may be one of the most important factors that has led to the diversity of specialized metabolism that we see in extant plants [17]. Mutations in enzymes involved in primary metabolism often result in significant loss of fitness or lethality. Moreover, primary metabolism is ubiquitous whereas specialized metabolism is lineage-specific; as such these lineage-specific biochemical routes are much younger than those of the primary metabolism. It is thus to be expected that primary metabolism experiences stronger evolutionary constraints than the specialized metabolism. The observed diversity of routes of the specialized metabolism may also be explained by an additional factor. Not only is it that—as discussed above—metabolic noise can become a functioning signal because it allows for higher fitness in a given environment, but drift can (and will) also play its part. Mutations in enzymes of the specialized metabolism that alter substrate specificity may be less likely to be lethal compared to the primary metabolism, thus if an ancestral enzyme is highly promiscuous the chances for lineage-dependent metabolic promiscuity is high. The phenylpropanoid pathway, although the evolutionary forces behind are still to be investigated, is a point in case for such lineage-specific promiscuity.

The phenylpropanoid pathway, and the metabolic routes that arise from this pathway, are at the heart of plant biology. From there derives lignin, which is essential for all vascular plants as a structural component for, e.g., erect growth, xylem vessel formation, and structural reinforcement upon pathogen attack (e.g., [33–35]). Additionally, derivatives of the phenylpropanoids can act in the responses to biotic and abiotic stress [36]. These routes—as can be derived from the name ‘specialized metabolism’—more often than not produce lineage-specific metabolites [17]. The chemodiversity of the phenylpropanoid pathway is underpinned by often large enzyme families that act within this pathway—with multiple steps being carried out by the same/similar enzymes acting on different compounds (see discussions in: [37–39]). Most of the genetic and molecular data that disentangle the role of the enzymes producing these metabolites comes from few select systems such as *Arabidopsis thaliana* and scattered data from other species. Given that some routes produce lineage-specific compounds, deriving interpretations from hard data of only a few organisms can be challenging. That said, current phylogenomic and transcriptomic data allow us to (i) gather a rough idea on how similar the phenylpropanoid pathway and its derived metabolism is between land plants and streptophyte algae and (ii) understand how the phenylpropanoid pathway evolved.

Earlier phylogenetic data suggested that phenylalanine ammonia-lyase (PAL)—the first enzyme in the phenylpropanoid pathway (Fig. 1)—occurred first in land plants and was possibly gained through horizontal gene transfer (HGT; [40]). Moreover, a gene coding for cinnamate-4-hydroxylase (C4H), the second enzyme in the pathway (Fig. 1), is missing in the genomes of green algae [39,41,42]. These observations stood in contrast to the presence of lignin-like compounds in streptophyte algae and the product of C4H in several different algae [43–47]. This conundrum hinted that other entry points to the phenylpropanoid pathway or other enzymes that convergently evolved PAL

and C4H activity might exist. Indeed, the availability of the first streptophyte algal genome—that of *Klebsormidium nitens* [48]—revealed the presence of a truncated version of a PAL-encoding gene [39,42]. Yet, its functions still need to be elucidated.

The finding of a PAL candidate in *K. nitens* supported that the phenylpropanoid pathway may have been present already in streptophyte algae. The key role of the phenylpropanoid pathway in biotic and abiotic stress (see [36]) led to the hypothesis that a version of the phenylpropanoid pathway was not only present in, but also helpful for the algal lineage that conquered the terrestrial habitat [8,29,42]. Indeed, putative CAD-encoding homologs responded to temperature stress on a transcript level in Zygnematophyceae [49]. Genome data of the Zygnematophyceae—the streptophyte algal lineage closest to land plants [50]—*Penium margaritaceum* yielded surprises as no PAL homolog could be identified [51]; despite that, *P. margaritaceum* produces flavonoids [51], which supports the notion that the core phenylpropanoid biosynthesis pathway functions in some fashion in this zygnematophycean alga.

How is phenylpropanoid biosynthesis realized in streptophyte algae that lack PAL? Possible options are (i) the evolution of PAL-function in phylogenetically related enzymes (such as histidine ammonia-lyases [HALs]) or (ii) (phenylalanine)/tyrosine ammonia-lyase ([P]TAL) as an alternative entry point to the phenylpropanoid pathway. PTAL converts tyrosine to coumarate and is known to occur in some land plants [52,53]. Further, cases of convergent evolution of enzymatic functions in the phenylpropanoid pathway have been identified for the function of F5H and COMT in *Selaginella moellendorffii* [54–56]. Both hypotheses are hence theoretically equally likely. That said, canonical homologs for ammonia lyases (PAL and HAL) have only been identified in *K. nitens* and *Chara braunii* [39]; albeit very short versions and tRNA-fusions of ammonia-lyases also exist in other streptophyte algal lineages [39]. This pattern is mirrored across the evolution of the phenylpropanoid pathway. One-to-one orthologs rarely exist in streptophyte algae for any embryophytic enzyme in the phenylpropanoid pathway and routes towards lignin [39]. Rather homologs and co-orthologs to many of the enzymes were found (Fig. 1, Fig. 2). The evolution of the enzymes involved in the biosynthesis of phenylpropanoids and lignin-like compounds is best depicted by radiation and lineage-specific duplications [39]. The presence of phenylpropanoid derivatives in representatives of the streptophyte algal lineages may thus derive from lineage-specific solutions. This is also supported by a phylogenomic analyses on the flavonoid pathway across streptophytes [57]. What exact toolkit may have been present in the streptophyte algal lineages that was the foundation of the terrestrial flora is however hard to infer.

There are several examples across streptophyte algae that suggest the presence of phenylpropanoid-associated and phenylpropanoid-derived compounds; many of these might be linked to stress responses. Among others, the accurate and fast molecular response to UV-irradiation is crucial for plant survival and thus the presence of UV-absorbing compounds, such as phenylpropanoid pathway-derived anthocyanins and flavones, is advantageous. The latter was not only detected in the zygnematophycean alga *P. margaritaceum* [51], but also in different species of chlorophyte algae ([47], Fig. 1, Fig. 3). Additionally, several studies identified phenolic compounds in zygnematophycean algae that act as sunscreens to protect against UV-irradiance [58–60]. For example, the alga *Zygonium ericetorum* does not only produce photoprotective phenolics but also hydrolyzable tannins [58]; tannins are known from land plants to be involved in various abiotic and biotic stress response mechanisms [61]. In the newly discovered species *Serritaenia testaceovaginata* (formerly belonging to genus *Mesotaenium* Nägeli), the formation of a pigmented extracellular mucilage provides photoprotection [62]. Under certain environmental conditions [63] some streptophyte algae form Mycosporine-like amino acids (MAAs) as a form of protection against UV-irradiance [64,65]. MAAs derive from the shikimate pathway and are based on the precursor 3-dehydroquinic acid; these amino-acid derivatives can absorb UV-light (Fig. 1; [65]). MAAs were found in several klebsormidiophycean algae, for example, the alpine

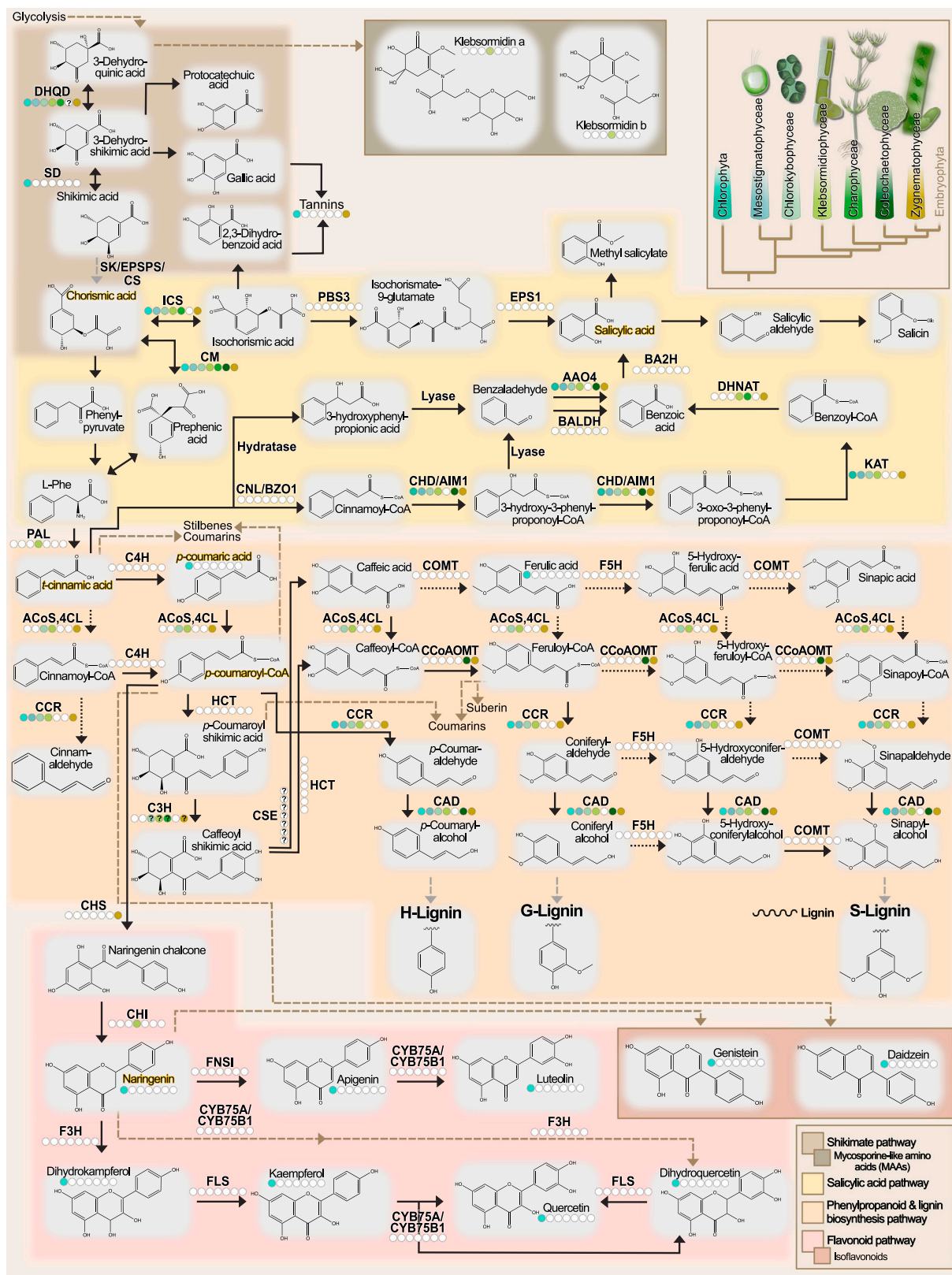


Fig. 1. Selected biosynthetic routes towards plant phenolics. An overall schematic interconnective pathway map for some possible biosynthetic routes towards plant phenolics. Enzymes, for which at least phylogenetic evidence is reported (for more, see Fig. 2) are colored in the corresponding color for a given lineage (see inset for the color key). Besides streptophyte algae, presence/absence patterns of enzymes and compounds of the Chlorophyta clade are shown. White spaces indicate no information about the presence/absence of the corresponding enzyme/compound. Question marks indicate that the presence/absence of the enzyme is inconclusive. Dashed arrows indicate steps in the pathway which are not shown or are putative/ambiguous. Highlighted in yellow are compounds which hold a key function in the shown pathways. The data are based on [39,47,58,65,99,305,306].

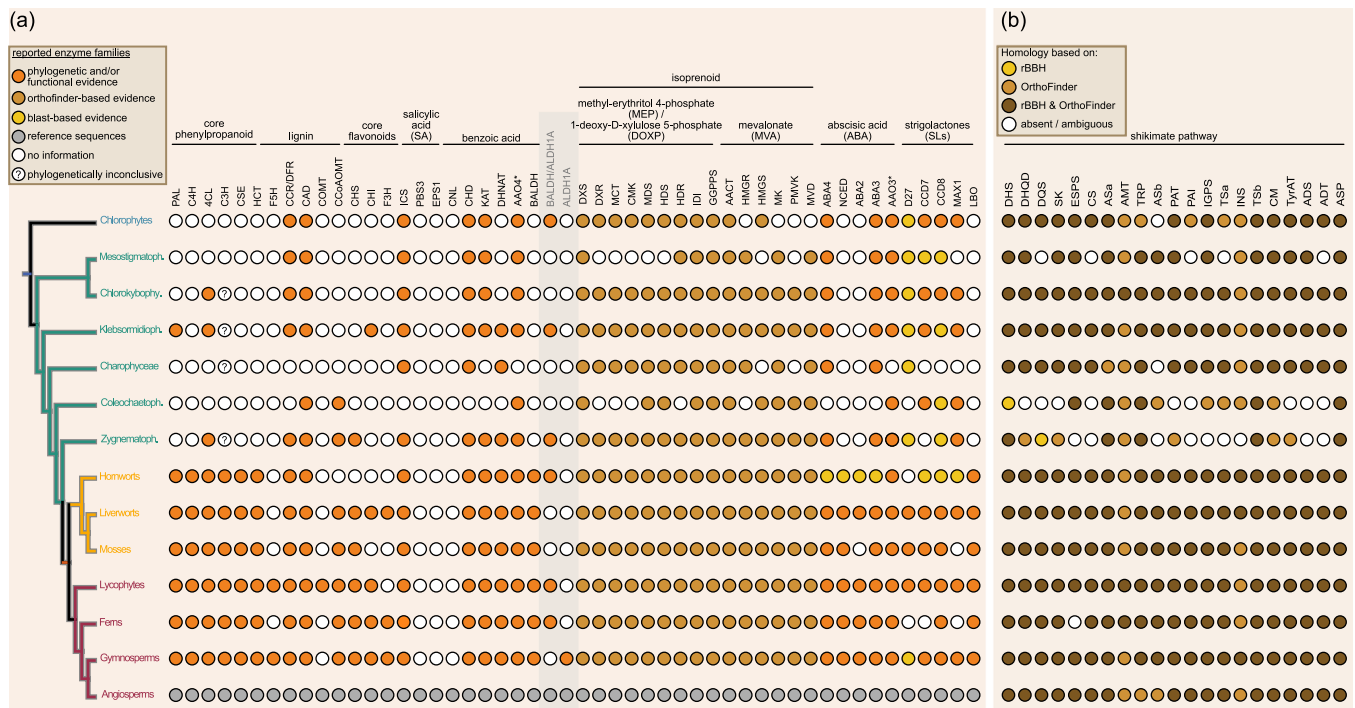


Fig. 2. Genetic repertoire for key specialized metabolic routes across streptophytes. (a) Evidence for the presence of gene families encoding enzymes involved in the synthesis of discussed specialized metabolites in streptophytes (for phenylpropanoids, lignin, core flavonoids, SA, benzoic acid, MEP/DOXP, MVA, ABA, SL). Colored dots represent reports on the presence of homologs belonging to the respective enzyme families based on: phylogenetic and/or functional evidence (orange), BLAST-derived conclusions (yellow), evidence from orthofinder (harvest gold). Question marks indicate inconclusive phylogenetic affiliation. White circles show that presence has not been reported for these lineages yet for any reason. Angiosperms provide the reference (grey dots). The data are based on [39,42,48,51,57,99,103,104,115,130,182,186,187,190,305–309]. The putative presence of homologous enzymes in the biosynthesis pathways of terpenoids and mevalonate was investigated here using Orthofinder2 using standard settings. (b) The putative presence of homologous enzymes involved the shikimate pathway was also investigated here using Orthofinder2. The list of enzymes involved in this pathway was derived from KEGG. All Orthofinder2 results are shown in Table S1 and the dataset was based on [48,51,102–104,114,115,130,307,310–337].

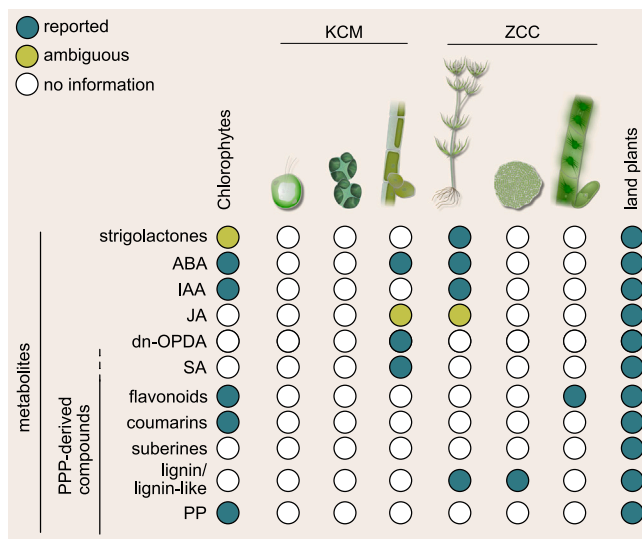


Fig. 3. Reports on important specialized metabolites in streptophyte algae. Streptophyte algae are listed in the order Mesostigmatophyceae, Chlorokybophyceae, and Klebsormidiophyceae (KCM); Charophyceae, Coleochaetophyceae, and Zygnematophyceae (ZCC). Reports on the presence (blue) of specialized metabolites, including phenylpropanoids (PP; PPP = phenylpropanoid pathway), its derivatives, cell wall components and phytohormones. The existence of contrasting reports is highlighted by yellow-green circles. White circles indicate that presence of these metabolites have not been reported (to our current knowledge) for a certain lineage for any reason. The data are based on [47,54,67,100,338–349].

biological-soil-crust-forming alga *Klebsormidium fluitans*, as well as several other *Klebsormidium* isolates and the species *Interfilum massjukiae* and *I. terricola*, which belong to the sister lineage to the genus *Klebsormidium* [64–66]. Klebsormidiophyceae as well as several other classes of streptophyte algae have homologs to the entire suite of enzymes involved in the shikimate pathway (Fig. 2b). Together, not only transcriptomic and phylogenomic data support the presence of a phenylpropanoid pathway version in streptophyte algae, but these studies add metabolomic evidence to the role of phenylpropanoid pathway-derived compounds in streptophyte algal stress responses.

As illustrated, a major challenge for terrestrial streptophyte algae is that they face high UV stress against which phenolic compounds provide shelter. Enrichment of the cell wall with phenolic compounds is one of the modes of building up UV protection, which is proposed for streptophyte algal phenolics [28]. Renault et al. [67] have shown that the bryophyte *Physcomitrium patens* enriches its cuticle with phenolics by using an enzymatic toolkit that is, in vascular plants, ascribed to lignin biosynthesis; streptophyte algae are known to produce a suite of compounds that form sporopollenin [43,68,69]. Additionally, recent data show that liverworts produce a novel compound termed auronidin, which is a novel class of anthocyanins, but derived from auronones [70]. This illustrates the option for as-of-yet unknown compounds involved in abiotic stress-responses. It is tempting to speculate that routes of the phenylpropanoid pathway shared by streptophyte algae and land plants trace their evolutionary roots to the production of an ancestral polymer [71] that formed a chassis for or direct aid in UV protection.

4. Salicylic acid: an ancient signal with two possible routes for its biosynthesis

Salicylic acid (SA) is a well-studied phytohormone in plant defense [72]. Chemically it is a derivative from benzoic acid; indeed, early studies on tobacco, potato, and cucumber highlighted PAL as a key enzyme in SA biosynthesis [73–75]. In *A. thaliana*, however, the main source of SA appears to be chorismate [76,77]. Isochorismate synthase 1 and 2 (ICS1, ICS2) catalyze the conversion of chorismate to isochorismate [76,78]. ENHANCED DISEASE SUSCEPTIBILITY 5 (EDS5) transports isochorismate from the chloroplast to the cytosol where GRETCHEN-HAGEN 3.12 (GH3.12 also known as PBS3) conjugates it to isochorismate-9-glutamate [79–81]. From there, SA can be derived by a spontaneous reaction in *A. thaliana* [80]. Additionally, Brassicaceae genomes encode a BAHD acyltransferase, named ENHANCED PSEUDOMONAS SUSCEPTIBILITY 1 (EPS1), that can facilitate the conversion from isochorismate-9-glutamate to SA [81,82]. This already hints towards some lineage-specificity in the biosynthesis of SA across land plants. In agreement with this notion, soybean appears to require both the ICS- and PAL-derived biosynthesis of SA for a proper response to pathogen attack [83]. In tobacco, SA accumulation is altered in leaves that were inoculated with tobacco mosaic virus (TMV) and systemic leaves when expression of PAL was reduced [74]. Yet, reduction of PAL in tobacco affected systemic acquired resistance (SAR) more strongly than the primary defense response [74].

Seminal work over the past few years has uncovered that, rather than SA alone, *N*-hydroxypipicolinic acid (NHP) and SA act in concert to induce SAR [84]. As such, the results by Pallas and colleagues [74] on SAR may need to be viewed in light of the recent findings. Additionally, Catinot and colleagues [85] found that an immediate induction of SA after pathogen attack requires ICS in *Nicotiana benthamiana*. In pepper (*Capsicum annuum*), silencing of *CaPAL1* decreased PAL activity as well as the production of SA and downstream SA-dependent expression of *PATHOGENESIS-RELATED PROTEIN 1* (*CaPR1*; [86]). Cumulatively, those studies highlight that the PAL- and ICS-pathways are required for SA production; although in which context and how much they contribute seems to vary by lineage—possibly by species.

The pathway by which SA is synthesized via PAL remains largely unknown. Nonetheless, several possible routes have been highlighted, in which several intermediate enzymatic steps have been functionally validated [87]. One option is that cytosolic *t*-cinnamic acid—derived from phenylalanine via PAL—is converted into benzaldehyde by a hydratase (possibly *PhCHD/AtAIM1*) and a lyase [87]. Alternatively, *t*-cinnamic acid could be converted first to its CoA-ester, cinnamoyl-CoA by 4CL and then processed as described above [87]. Cytosolic benzaldehyde would then either be converted to benzoic acid via Arabidopsis thaliana ALDEHYDE OXIDASE 4 (AAO4) in *A. thaliana* [88] or benzaldehyde is transferred into the mitochondrion, which is the case in *Antirrhinum majus* where benzaldehyde dehydrogenase (BALDH) catalyzes the reaction [89]. The second option would require the peroxisomal β -oxidative pathway and as such import of *t*-cinnamic acid or its CoA-ester to the peroxisome. In that part of the putative pathway the steps from *t*-cinnamic acid via cinnamoyl-CoA to benzoic acid have been functionally elucidated in either *Petunia hybrida* or *A. thaliana* [90–97]. Independent of where in the cell benzoic acid is synthesized, it is supposed to be converted into SA in the cytosol. León et al. [98] partially purified a benzoic acid-2-hydroxylase (BA2H) and showed that this CYP450 enzyme is capable of catalyzing this step. Similarly, experiments in pea have also worked on protein level with a BA2H-candidate. Yet, no sequence from BA2H has been published from any plant species. This is a major bottleneck in identifying the evolutionary conservation of all possible SA biosynthesis pathways.

The foremost question that allows for inferences on the evolution of SA biosynthesis, however, is: how conserved is the occurrence of SA across streptophyte diversity? Outside of angiosperms, SA has been measured, e.g., in the floating fern *Azolla filiculoides* (although only

small amounts; [99]), several bryophytes [100] and the streptophyte alga *K. nitens* [48]; Fig. 3). This suggests that SA was present in the last common ancestor (LCA) of most, if not all, streptophytes [9]. Deep homologs for certain genes of SA biosynthesis and signaling can even be found in chlorophytes [101,102]. The emergence of SA signaling and its role in defense is currently placed at the LCA of land plants [9,10]. More experimental data from streptophyte algae are needed to confirm this current placement. That said, the genomic insights into possible biosynthetic routes have already been informative.

In the genome of the Charophyceae *C. braunii* no homolog to ICS was recovered via a BLAST search ([103]; Fig. 2). This is in contrast to SA being found in *K. nitens* [48] that belongs to the Klebsormidiophyceae that branch sister to Phragmoplastophyta, which include *Chara*. Phylogenetic analyses on the SA biosynthesis pathway in three hornwort genome assemblies (*Anthoceros* spp.) not only uncovered ICS candidates for the hornworts but also *C. braunii* [104]. Yet, sometimes ICS may be confused with other enzymes bearing chorismate-binding domains. In a more recent dataset including more species as well as sequences of the other chorismate-binding enzymes anthranilate synthase, amino-deoxychorismate synthase/para-aminobenzoate synthase and chorismate synthase the affiliations to either group appear clearer: Indeed, ICS candidates occur throughout the entire green lineage—not only in *C. braunii* but also several chlorophytes [99]. This suggests a deep evolutionary origin of ICS. Concomitant a transfer from the cyanobacterial plastid progenitor to the nucleus via endosymbiotic gene transfer has previously been suggested [105]. Additionally, functional homologs of ICS are present in the genomes of several filamentous cyanobacteria [99,106]. The ICS-derived SA biosynthesis pathway is ancient and bacteria use either isochorismate pyruvate lyase for conversion of isochorismate into SA or salicylate synthase for conversion of chorismate to SA [107–109]. Alternatively, isochorismate can be funneled into the menaquinone pathway in bacteria [110,111]. And the presence of menC and/or menD domains attached to some ICS of land plants and algae [99] suggest that they may be rather involved in phyloquinone synthesis. Exceptions to the occurrence of these compound enzymes are the sequences of angiosperms, that of the lycophyte *S. moellendorffii* and that of *C. braunii* [99].

What about the biosynthesis pathway via PAL? As described above, the evolutionary history of PAL is complicated. While *K. nitens* clearly has a PAL-candidate, other streptophyte algae seem to lack PAL [39]. Likewise, the occurrence of 4CL homologs—required if cinnamoyl-CoA is the precursor—are only present in a few streptophyte algal lineages sequenced to date ([39]; Fig. 2). Other steps such as that carried out in the peroxisome by the hydratase *AtAIM1/PhCHD* or the last step to benzoic acid carried out by *DHNAT* are conserved across streptophytes ([39]; Fig. 2). Nonetheless, possible candidates for all steps in the cytosolic and peroxisomal route exist throughout the entire lineage of streptophytes (Fig. 2). This makes sense, because these routes are not entirely devoted to SA biosynthesis, but products are also required for the synthesis of other essential compounds. The only enzyme specific to SA synthesis from this pathway is BA2H—whose sequence currently remains an enigma. The thrive of non-seed model systems and the continued development of (streptophyte) algal model systems [48,103, 112–117] will be a crucial step to functionally determine the evolutionary routes of SA biosynthesis and its diversity.

While SA was likely present in the LCA of streptophytes, what role this phytohormone occupied in that LCA is not elucidated. As mentioned above—based on current data—its role in defense signaling can be traced back to the LCA of land plants [9]. The evolutionary origin of the cross-talk between SA and jasmonic acid (JA) to modulate different defense responses towards different types of pathogens can be placed either at the LCA of seed plants [118] or at the LCA of angiosperms [9]. Recent data by Matsui et al. [119] indicate that cross-talk between SA and jasmonates might be present in the liverwort *M. polymorpha* and thus possibly already in the LCA of land plants. Data from the moss *P. patens* [120] and the gymnosperm *P. abies* [121] indicate a role for SA

in response to necrotrophic pathogens; which canonically is ascribed to JA and regulated by the antagonistic relationship between SA and jasmonates. Thus, lineage-specific modulations of the antagonism could confound its origin.

SA is involved in the response to abiotic stresses as well (e.g. [122, 123]). Further, SA interferes with gravitropism by altering the flux of auxin in *A. thaliana* [124]. Cross-talk of SA with other hormones is an important modulator of the trade-offs between defense and growth or defense and response to abiotic stress. How conserved these interactions of phytohormones are—i.e. whether this can be found across the streptophytes—remains to be elucidated. A cornerstone in discussing the evolutionary origin, distribution, and composition of phytohormone networks is to map the occurrence of the key players onto the lineages of all extant streptophytes—including streptophyte algae. This is the starting point for any robust evolutionary inference. In the next sections we will thus address the presence of other phytohormones across extant streptophytes and try to infer what networks may have been theoretically possible in various LCAs of streptophyte algae and land plants.

5. Jasmonates: ancient roles, new functions, and shifts in ligand preferences

Oxylipins comprise and contribute to a plethora of different compounds. They are derived from unsaturated fatty acids via the oxidation of lipids, including derivatives of these molecules. As a phytohormone, the oxylipin jasmonic acid (JA) is widely studied in plant biology. It is important for the defense against herbivores and certain lifestyles of pathogens [125,126]. To achieve a fine-tuned response, JA signaling is tightly interwoven with other phytohormone signaling pathways—most prominent is the interaction with ethylene (ET) and its antagonism with SA [125,126]. Both, ET and SA have been detected in streptophyte algae [48,127] (more on ET below), and may thus have likely been present in the earliest land plants. In contrast, JA likely emerged later during the evolution of land plants [128,129]. Concomitantly, the genomes of sequenced streptophyte algae reveal an absence of several of the JA biosynthesis genes [48,51,103,130]. However, allene oxide synthase (AOS, which catalyzes the conversion from 13-HPOT to 12,13-EOT, an intermediate to JA, in angiosperms) was identified and functionally characterized in *K. nitens* (then still called *K. flaccidum*), *P. patens* and *Marchantia polymorpha* [131,132].

Functional characterization of allene oxide cyclase (AOC) is also available for *P. patens*, which does not produce JA [133]. Moreover, Pratiwi and colleagues [134] functionally characterized four enzymes in the JA biosynthesis pathway (AOS, AOC, OPR3, and JAR1) in *S. moellendorffii*; JAR1 produces the functional JA conjugate, JA-Ile [135]. In *S. moellendorffii*, the canonical function for all four enzymes has been reported [134]. However, in *P. patens* AOCs cannot only produce *cis*-(+)-OPDA, but also 11-OPDA [133]. Together, the patchy occurrences—or possible lack—of JA in bryophytes and streptophyte algae [48,132,133,136,137] (Fig. 3) as well as the functional promiscuity of PpAOC2 [133] indicate that JA signaling may not be conserved across streptophytes. Monte and colleagues [128,129] investigated this aspect of JA signaling in *M. polymorpha*. Indeed, they found that the JA-Ile receptor and the entire downstream signaling cascade leading to JA-derived responses are conserved in *A. thaliana* and *M. polymorpha* [128,129,138]. There was just one difference: the ligand of the JA-Ile receptor (CORONATINE-INSENSITIVE 1; COI1) of *M. polymorpha* recognized dn-OPDA, not JA-Ile [128]. The authors of that study [128] found that the shift from dn-OPDA to JA-Ile was attributable to a mutation in the COI1 sequence; this mutation leading to JA-Ile being the preferred ligand might date back to an older ancestor, as it was in several angiosperms, *Amborella trichocarpa* (which belongs to the sister lineage to all other angiosperms) as well as the lycophyte *S. moellendorffii*.

The role of dn-OPDA in the liverwort *M. polymorpha* partially overlaps with that of *A. thaliana* [128]; being relevant for defense against

herbivory. Yet, an effect on fertility, as observed for *A. thaliana*, was not found for *Mpcoi1* [128]. In the moss *P. patens* only PpAOC2, but neither PpAOC1 nor PpAOS1 nor PpAOS2 affected fertility [131,133]. This begged the question of what the ancestral role of the dn-OPDA/JA-Ile pathway was. Studies on the *Mpcoi1* mutant suggested the presence of a COI1-independent signaling of dn-OPDA [139]. This pathway transcriptionally induced many responses related to abiotic stresses, and mostly temperature stress. Monte and colleagues [139] were able to confirm this independent signaling pathway of dn-OPDA and could show that this is conserved in *A. thaliana* and the alga *K. nitens*. Together with the observation that *K. nitens* as well as other streptophyte algae may lack a COI1 homolog [48,103,115,130,140], this might point to an ancestral role of dn-OPDA in thermotolerance via a COI1-independent pathway [139].

6. Apocarotenoids: key signals under stress that emerge from essential pigments

6.1. Isoprenoid building blocks for essential plant functions

Isoprenoids, which include the diverse terpenoids, are a large class of organic molecules. Terpenoids include carotenoids, phytols, retinols, tocopherols, dolichols, and squalene; also, chlorophylls, cytokinins, and brassinosteroids are among the isoprenoid-derived molecules. And this list of important molecules could be continued. As such, they are critical to the molecular biology of plants and algae; changes in terpenoid biosynthesis affect growth, development, and stress responses of embryophytes through fine-tuning of phytohormone production [141]. Terpenoids are classified based on the number of isoprene units or the type and number of cyclic structures. They are produced from the condensation of dimethylallyl pyrophosphate (DMAPP) and its isomer isopentenyl pyrophosphate (IPP) [142]. Among the various terpenoids, archaea, many eukaryotes such as animals, and some gram-positive bacteria employ the mevalonate (MVA)-pathway exclusively, while gram-negative bacteria, some gram-positive bacteria, and—importantly—chlorophyte algae *only* employ the methyl-erythritol 4-phosphate (MEP)-pathway [141,143]. Embryophytes use both the cytosolic MVA-pathway and the plastidic MEP-pathway that was brought into photosynthetic eukaryotes through the cyanobacterial plastid progenitor [141,143]. Given the paraphyletic relationships within the streptophytes, there looms the question of when, during the course of Chloroplastida evolution, the retention of MEP and MVA or the shift solely towards MEP occurred.

Using Orthofinder2 [144], we looked for MEP and MVA enzymes (based on *A. thaliana* gene IDs) in orthogroups formed by phylodiverse protein data from across the green lineage (Fig. 2). If there was at least one species in a lineage that had a homolog to the *A. thaliana* gene, we considered that this lineage has a homolog of the gene. Our survey suggests that the MVA pathway is present completely in some streptophyte algae (Chlorokybophyceae, Klebsormidiophyceae, and Zygnematophyceae) and partially in all of them. Importantly, this includes the first committed step carried out by 3-hydroxy-methyl-glutaryl-CoA reductase (HMGR), which is present in all embryophytes and some streptophyte algae including *Chlorokybus*, *K. nitens*, *C. braunii*, *Spiroglaea muscicola*, and *Mesotaenium endlicherianum*. This is in line with Pu et al. [141], who reported that all enzymes of the MEP pathway are present in embryophytes, streptophyte algae, and Chlorophyta (except for *Mesostigma viride*, *S. muscicola*, and *M. endlicherianum*); the only difference is that, based on our analyses, Chlorokybophyceae also have the complete MVA pathway. Pu et al. [141] speculated that emergence and/or retention of the MVA pathway in embryophytes may facilitate the transition from shallow freshwater habitats to subaerial and terrestrial environments. Indeed, given the importance of isoprenoid-derived compounds for the acclimation to environmental conditions, it is conceivable that a concerted action of MVA and MEP offers spatio-temporal versatility in responses. For example, Pu et al. [141]

highlight that *K. nitens*, which grows in terrestrial environments, has the genes for the entire MVA-pathway. In contrast, a unicellular green alga that thrive in shallow ponds, *M. viride*, has only four MVA-pathway genes. In accordance with this speculation, aquatic Chlorophyta have only one (AACT) or two (AACT and HMGS) enzymes of MVA-pathway. The distribution of the MVA versus MEP pathway thus remains obscure. What is however clear is that streptophyte algal isoprenoid biosynthesis (very likely via MEP) leads to carotenoids.

6.2. Carotenoids and xanthophylls: first responders to stress in photosynthetic eukaryotes

Carotenoids are a class of specialized metabolites that occur in all photosynthetic eukaryotes and cyanobacteria [145]. Due to the central role as accessory pigments in photosynthesis, carotenoids are essential for plant physiology. That said, some carotenoids are among the classes of specialized metabolites that are rare and often species-specific [145]. Even though the biosynthesis of carotenoids is well-conserved across the green lineage, this does not mean that carotenoids are mere steady-state essentials: carotenoids are critical in the response to many terrestrial stressors that land plants face.

Among the first cellular machineries that respond to adverse environmental conditions are the photosynthetic apparatuses. This also holds true for streptophyte algae closely related to land plants [49, 146–149]. Especially upon the rapidly changing light conditions in the terrestrial habitat. A key mechanism for mitigating excess energy is non-photochemical quenching (NPQ; [150,151]). While there are important differences in NPQ mechanisms across the diversity of the green lineage [152–154], many key NPQ-triggering mechanisms such as the action of light-harvesting complex stress-related protein (LHCGR) and photosystem II subunit S (PSBS) [155–159] are conserved (see [8] for a longer discussion). One of the main processes tied to NPQ is the xanthophyll cycle and thus carotenoids.

At high photosynthetic rates, the lumen acidifies and the conversion of the xanthophylls violaxanthin to zeaxanthin is triggered [160,161]. This conversion is intertwined with NPQ [162,163]. Such conversions hence hinge on the quantity and composition of the xanthophyll/carotenoid pool. Only few studies so far investigated the carotenoid profiles of streptophyte algae. For example, Stamenković and colleagues [164] measured carotenoid and chlorophyll composition in *Cosmarium* isolates from different geographic areas under high light stress. They found that the investigated desmids (with exception of the arctic isolate *C. crenatum* var. *boldtianum*) showed pigment characteristic for a complete xanthophyll cycle as in land plants that are subjected to high irradiance [164]. Given that desmids are primarily found in shallow freshwater habitats the authors suggested that the observed increase of the xanthophyll pool and its fast action of the xanthophyll cycle might help desmids to cope with the relatively high light intensities and fluctuations in their natural habitat [164]. Importantly, such a rich pigment pool adds to the versatility of specialized metabolism as they act as a substrate for more metabolites.

Carotenoids are involved in the mitigation of one of the foremost consequences of terrestrial stressors: reactive oxygen species (ROS) [145,165,166]. By quenching ROS, a class of by-products is formed that has been shown to be used extensively in embryophyte stress response—the apocarotenoids [166–168]. It is only conceivable that during the course of evolution, by-products that increase in abundance under stress turn into signaling molecules that trigger regulatory cascades for stress response.

6.3. Apocarotenoids: cleavage products key in stress response

Apocarotenoids include some of the best-known signals of specialized metabolism in embryophytes: examples are the phytohormones abscisic acid (ABA) and strigolactones (SL), and well-accepted retrograde signaling molecules such as β -cyclocitral (β -CC) [167–169].

Almost every aspect of plant biology is influenced by apocarotenoids (for an overview, see [145,168]). Apocarotenoids arise through oxidative cleavage of carotenoids either in a non-enzymatic fashion (e.g., through ROS) or via the action of Carotenoid Cleave Dioxygenases (CCD). Apocarotenoids are among those signature molecules involved in the embryophytic response towards terrestrial stressors, including the prime stressors high irradiance and drought [166,168]. It is thus conceivable that these molecules were among the foundational specialized metabolites that aided in the conquest land.

Apocarotenoids are formed via spontaneous degradation of the evolutionarily conserved carotenoids [145,166–168]. Thus, from a chemical perspective their presence would be expected in streptophyte algae. Indeed, Hori et al. [48] detected ABA in *K. nitens* alongside genes putatively coding for parts of the ABA biosynthesis and signaling pathways. Yet, homologs of ABA receptors of the PYRABACTIN RESISTANCE family (PYR/PYL/RCAR) are missing in *K. nitens*; also no clear ortholog of the biosynthesis-relevant 9-cis-epoxycarotenoid dioxygenase (NCED) were identified [48]. Recent investigations into the ABA biosynthesis of *A. thaliana* have revealed an alternative route via diverse β -apo-11-carotenals [170]. This (a) illustrates that also in well-researched plants there are novel biochemical routes to be found and (b) opens a whole range of possibilities that might provide an explanation for the occurrence of ABA in streptophyte algae that lack NCEDs and ABA DEFICIENT 2 (ABA2) homologs (see, e.g., [103,149]). In contrast to *Klebsormidium*, the Zygnematophyceae have genes orthologous to those coding for a PYL receptor in land plants (Fig. 4) [115,149]. Moreover, the zygnematophycean PYL-like protein showed basal inhibition of its downstream-targeted phosphatases (PP2C)—similar to land plants—but did not show an increased inhibitory activity upon ABA treatment [171]. Hence, ABA signaling, as known from land plants, differs in streptophyte algae (see also [8,172]). The possibility, however, remains that the zygnematophycean pre-PYL receptor—in a similar scenario as shown for the JA receptor [128]—can bind a different specialized metabolite

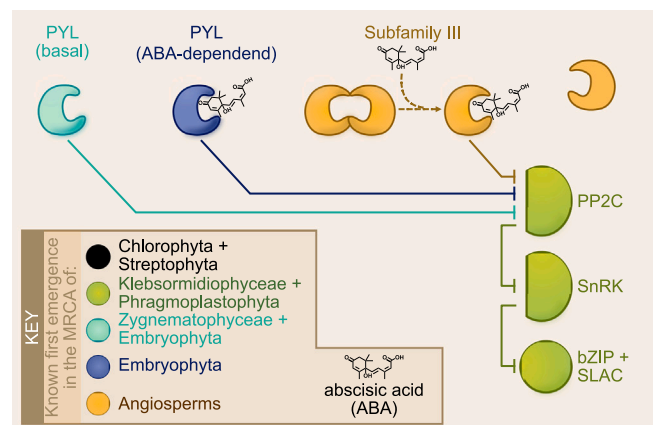


Fig. 4. Modular evolutionary origin of the abscisic acid signaling cascade. ABA biosynthesis likely originated in the last common ancestor (LCA) of Chlorophyta and Streptophyta (hormone in black)—since ABA has been detected in several chlorophytic algae, streptophytic algae (see Fig. 3 and main text). The bona fide signaling pathway has been elucidated in embryophytes (see [350] for a review): ABA perception and core signaling is based on PYL, PP2C, SnRK2s, transcription factors (basic leucine zipper, bZIP), and ion channels (S-type anion channels, SLAC) [351–353]; the regulatory connection between PP2C, SnRK2s, and transcription factors/ion channels likely already emerged in Klebsormidiophyceae [171,172,352,353]. Hence, this cascade likely emerged in the LCA of Klebsormidiophyceae and Phragmoplastophyta (green)—potentially even earlier. The mechanism of basal inhibition by a ligand independent PYL was gained at the base of Zygnematophyceae and embryophyta (cyan) [149,171,172]. The ligand-dependent PYL inhibition of the signaling network is land plant specific (blue) and the dimeric PYLs that monomerize in an ABA-dependent manner are only known in angiosperms (subfamily III; yellow) [171].

(structurally related or not) than in Embryophyta.

SLs are apocarotenoids that originate from β -carotene and represent an important class in the specialized metabolism of embryophytes [168, 173]. The SL signaling network is involved in symbioses with arbuscular mycorrhizal fungi (AMF) as well as many root and rhizoid modulating processes [168, 173–179]. It is prudent to note that next to SLs, other apocarotenoids mediate symbiosis with AMF; these include mycoradecin [180] and several blumenols [181]. Because SLs were reported in algae of the order Charales and in land plants that lost their ability to form symbioses with AMF [168, 182, 183], their ancestral function was considered to rather be associated with developmental processes than symbiotic signaling—but more on this below. Indeed, Delaux et al. [182] were able to induce the rhizoid elongation of *Chara corallina* (a Charophyceae) with a similar efficiency compared to those observed for embryophytes treated with a synthetic analog of SLs, GR24.

The enzymes CCD7 and CCD8 are crucial for SL biosynthesis in angiosperms [184] and mosses [185, 186]; CCD8 of the liverwort *Marchantia paleacea* was recently shown to act in SL biosynthesis [187]. Even though *Chara corallina* responds to SL application, Charophyceae appear to lack genes encoding for CCD7 and CCD8 [103, 188]. Likewise, the liverwort *Marchantia polymorpha* lacks homologs to CCD8 [182]; that said, Kodama et al. [187] recently showed that several *Marchantia* spp.—except *M. polymorpha*—encode both CCD8 and MORE AXILLARY GROWTH 1 (MAX1; [51]). The latter is a cytochrome P450 enzyme (CYP711A) relevant for the formation of carlactonic acid, from which multiple strigolactones arise [189]. Moreover, Kodama et al. [187] highlight that all species of *Marchantia* capable of engaging in symbiosis with AMF have the genetic repertoire to produce SLs. Additionally, the same study confirmed the presence of carlactonic acid as well as a new SL, bryosymbiol, in *Marchantia paleacea*. A recent phylogenetic analysis by Walker et al. [190] reported the presence of deep homologs for CCD7 and CCD8 in chlorophytes; albeit pointing out that these possible homologs appear rather divergent from land plant sequences, calling into question the conserved substrate specificity. Overall, homologs of both CCD7 and CCD8 occur across the green lineage (Fig. 2a). Enzymes with *bona fide* CCD7 and CCD8 activity were at least already present in the LCA of land plants; whether homologs of CCD7 and CCD8 in, for example, streptophyte algae encode enzymes with a similar function as described for their functionally characterized counterparts in embryophytes needs to be elucidated. Given that homologs of key components of SL biosynthesis have been reported to be present in most streptophyte algal genomes sequenced to date (with the exception of LBO; Fig. 2a), it would not be surprising if they could act in a similar biosynthetic chassis—with the theoretical capacity to give rise to a plethora of apocarotenoids. The detection of SLs outside of land plants has however been patchy (Fig. 3). Similar to the genetic distribution of the SL biosynthesis pathway, components in SL signaling are present in streptophyte algae; the full chassis of SL perception and signaling—as described from land plants—more likely came together in the last common ancestor of land plants [187, 191]. This stepwise emergence of the embryophytic SL biosynthesis followed by the emergence of the signaling chassis was coined the ‘ligand first’ scenario by Kodama and colleagues [187]. In turn this however means, that if there are ‘ligands first’, there are biosynthetic routes first. Such ligands emerge from a pathway with multiple steps and compounds—all of which might play important roles in the respective systems. Aided by emerging algal model systems, we can hunt for their role in cell signaling and physiology.

Several small volatile apocarotenoids, such as β -CC, β -ionone and dihydroactinidiolide, arise as products of CCD cleavage and by-products after spontaneous degradation of carotenoids upon oxidative stress [145, 166–168, 192, 193]. Also, these small apocarotenoids have diverse functions in physiology and beyond. If you enjoy a glass of red wine while reading this article, part of the aroma originates from β -ionone [194–198]. Many volatile apocarotenoids are well-known as herbivore repellents [199, 200]. The appearance of such small apocarotenoids is

known from diverse organisms, ranging from cyanobacteria to land plants [166–168, 201, 202]. Thus, even though no data are available from streptophyte algae, one can assume that they may also occur there. Indeed, there might be a lot of added value to studying other apocarotenoids than the usual suspects like ABA and SL in streptophyte algae. And there are a lot of possible candidates to pick. Over 750 different carotenoids [145] are currently known. Taking into account that apocarotenoids can arise from cleavage of carotenoids in nearly any position of the polyene backbone—either enzymatically or non-enzymatically—there are 10,000 s of theoretically possible apocarotenoids. Given this diversity, there is much we do not know about the function of a major share of apocarotenoid diversity.

β -CC has attracted some attention in recent years [167, 203, 204]. Treatment with β -CC elicited transcriptomic profiles that resembled those triggered by the ROS singlet oxygen [167], highlighting the whole cascade from (i) ROS as the input to (ii) β -CC generation and signaling to (iii) transcriptomic output. In addition to its possible role as a ROS sensor, β -CC has been shown to control the upstream isoprenoid pathway by acting as an inhibitor of 1-deoxy-D-xylulose 5-phosphate synthase (DXS), the first step in isoprenoid biosynthesis via the MEP pathway, in *A. thaliana* [204]. Hence, β -CC can be considered a potent and versatile regulator. β -CC gives rise to a spontaneous oxidation product, β -cyclogeranic acid (often also called β -cyclocitric acid, which can cause a bit of confusion as it is not the cyclic form of citric acid); β -cyclogeranic acid is involved in the response to drought stress [203]. Similarly, the spontaneous oxidation product of β -ionone, dihydroactinidiolide, is induced by high light and triggers the expression of singlet oxygen quenchers [193]. β -ionone itself is induced in a light-dependent manner, as work on the model system *Petunia hybrida* has shown [205]. If oxidative cleavage on a carotenoid happens two times, the middle section remains—a dialdehyde (DIAL). The function of this apocarotenoid class is one of the least studied—even though they were detected in cyanobacteria and land plants [168, 206]. One of these apocarotenoids is anchorene; already from the name its function can be guessed—anchorene is involved in the formation of anchor roots in *A. thaliana* [168, 207].

In sum, next to well-known apocarotenoid phytohormones, there is a series of apocarotenoid molecules with pronounced impact on plant physiology—many of which we could not cover here. Considering the potential and measurable diversity of apocarotenoids within one species, there is an astounding diversity of apocarotenoid metabolites across the green lineage. This will likely be matched by a diversity in apocarotenoid metabolism and function. It is likely that among that chemodiversity, evolution has recruited apocarotenoids with diverse roles in stress response, defense and signaling multiple times independently; the biological impact a given apocarotenoid might have in one lineage might not necessarily be the same in another. The impact of apocarotenoid chemodiversity is however writ large across the biodiversity of the green lineage.

7. Other important signals: glimpses into ethylene and auxin signaling in streptophyte algae

Auxin is a major growth hormone of land plants. Ultimately, auxin-mediated signaling influences almost any imaginable developmental and morphogenetic process of land plants. The *bona fide* source for auxin is the amino acid tryptophan. TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS (TAA) converts tryptophan into indole-3-pyruvic acid [208], which is further converted by flavin monooxygenases of the YUCCA family into indole-3-acetic acid (IAA) [209, 210]. While first described in angiosperms, this system was found to act in bryophytes such as *P. patens* and *M. polymorpha* [211, 212]. That said, there are some exceptions even in vascular plants; Kaneko et al. [213] described differences in auxin catabolism of the lycophyte *S. moellendorffii* leading to pronounced differences in the ratios of bioactive and inactive auxin as compared to flowering plants. The situation in streptophyte algae is

more complicated. While clear YUCCA orthologs have been reported for streptophyte algae [214], the emergence of *bona fide* TAA-coding genes is being debated [215,216]. Nonetheless, auxin has been detected in streptophyte algae [48,136]. Presence of auxin in streptophyte algae is, however, not surprising—auxin is even found in distantly-related algae as well as bacteria [217–219].

The action of auxin as a morphogenetic signal hinges on the creation of a spatial concentration gradient. This spatial auxin distribution is established by polar auxin transport—foremost orchestrated by the PIN auxin efflux carriers [220–222]. Auxin transport has been described in streptophyte algae. Boot et al. [223] found transport of IAA between the internodal cells of *Chara corallina*. Skokan et al. [224] carried out a detailed characterization of the single PIN homolog found in *K. nitens* (“KfPIN”). KfPIN transported auxin and was found (through immunolocalization) to localize at the surface of *Klebsormidium* cells [224]. That said, they did not localize at the contact points between cells of the filament [224]. Further, when Zhang et al. [225] heterologously introduced KfPIN into *Arabidopsis*, it localized to the plasma membrane in an even manner, thus, failing to establish polarity and a directionality in the flow of auxin. In contrast, a PIN from the Charophyceae *Chara vulgaris*, which shows erect growth and produces rhizoid, was found to localize in a polar fashion [226]. Moreover, genome analyses of another species from the genus *Chara*, *Chara braunii*, uncovered an expanded battery of PINs [103]. How polar auxin flow evolved (or how many times), and the likely roles that it played in the earliest land plants will keep plant scientists busy for decades to come.

Auxin acts through a signaling cascade that consists of a chain of negative regulation. At low auxin levels, the AUX/IAA proteins keep transcription factors of the AUXIN RESPONSE FACTORS (ARFs) family inactive [227,228]. Once auxin levels rise, the ubiquitin-ligase complex SCF^{TIR1} initiates ubiquitination and 26 S proteasomal degradation of the AUX/IAA repressors [229], releasing the ARFs that activate the auxin-dependent gene expression responses. ARFs come in three flavors: A, B, and C. They likely emerged at the base of streptophytes as ARFs of the class C, and somewhere along the trajectory of phragmoplastophyte evolution, the class A/B emerged [103,230–232]. However, since the *bona fide* SCF^{TIR1} complex is missing, these likely act in a molecular function that is independent of auxin signaling [231,232].

Ethylene is one of the few (if not the only) phytohormone for which the situation in streptophyte algae appears quite straightforward. The signaling cascade seems functionally conserved between land plants and Zygnematophyceae [127,233,234]. Exogenous application of ethylene on the zygnematophycean alga *Spirogyra pratensis* triggers the differential expression of stress- and photosynthesis-associated genes known from land plants [233]. This warrants particular attention as *S. pratensis* has more than just the full homologous chassis for ethylene signaling—*Spirogyra* homologs can functionally complement *Arabidopsis* mutants deficient in the respective ethylene signaling components [127]. Thus, ethylene (a) brings stress responses and (b) phytohormone-mediated signaling full circle—a role in stress response that is conserved between land plants and Zygnematophyceae.

8. Tinkering of pathways, by-products, and the evolution of specialized metabolism

8.1. On tinkering

How did the biosynthetic routes that underpin the chemodiversity of land plants emerge and radiate? A major source of innovation in evolution is gene duplication, as it provides opportunities to explore new gene functions while maintaining the original enzymatic activity [235, 236]. A duplicated gene can acquire mutations in its regulatory region leading to a variation in the time or tissue in which it is expressed (subfunctionalization), or mutations can affect the active site or the substrate-binding sites conferring a new function that differs from that of the ancestral gene copy (neofunctionalization). Protein domains of

duplicated genes can also combine in different ways to produce new proteins with new enzymatic capabilities. Gene and genome duplications have been relatively frequent during plant evolution [237] and are a major source of genomic innovation, including the plant metabolism [238]. The role of genetic redundancy contributing to metabolic complexity has deep roots in plant evolution, since their core metabolism was formed by enzymes derived from three genetic compartments (nucleus, mitochondrion, plastid) that ancestrally possessed many of the core metabolic pathways. This metabolic redundancy, along with genes acquired by endosymbiotic gene transfers (especially from the cyanobacterial plastid progenitor), have produced mosaic pathways with enzymes of diverse evolutionary origins [18]. In addition to providing new material for genetic innovation, redundancy has been proposed to sometimes confer metabolic robustness against environmental adversity, as an example see the multiple copies of the enzymes of the tricarboxylic acid pathway retained by many plant species [18].

Genetic redundancy allows existing enzymes (either from the core or specialized metabolisms) to be co-opted or hijacked into new specialized metabolic pathways. The re-use of existing material not only refers to full enzymes; their building blocks (protein domains) have also been combined in new different ways to perform a variety of enzymatic functions. Most protein domains found in enzymes of the specialized metabolism have deep evolutionary roots in primary metabolism [17]. For example, triterpenoid metabolism co-opted a component of the cell wall metabolism machinery via recruitment of cellulose synthase-like enzymes alongside other enzymes of saponin biosynthesis to the endoplasmic reticulum [239]. Other examples are the aromatization of most specialized metabolites, which evolved via specialization of the detoxification enzymes glyoxalases [240] or the biosynthesis of caffeine that evolved multiple times by co-option of similar ancestral enzymes [241]. Sometimes, new enzymes also evolved from non-catalytic ancestors: chalcone isomerase, one of the core flavonoid biosynthesis enzymes was recruited from fatty acid binding protein [242]. Depending on the source material and evolutionary fate of gene duplicates, evolution can lead to the same or diverging functions. In the presence of similar selective forces, similar metabolic pathways might evolve multiple times independently, either from the same ancestral protein domain (parallel evolution) or from different protein folds (convergent evolution). For example, *Selaginella* and angiosperms recruited paralogous OMT-encoding genes in the syringil (S) lignin biosynthesis [54] (which potentially might hold true—at least for some enzyme homologs—for other lycophytes [243]) and the conversion of flavonones to flavones is performed by the FNS II enzyme of the P450 family in most angiosperms and the FNS I enzyme of the ODD family in Apiaceae [17, 244]. More frequently, an ancestral protein might diverge in time and evolve different catalytic residues and substrate specificities (e.g. OMT evolved into Chalcone O-methyltransferase (ChOMT) and isoflavone O-methyltransferase (IOMT) in *Medicago*; [245]).

The fact that natural selection refines enzymatic function in specific cellular contexts speaks of the contingent nature of evolution: an enzymatic function will never be universally “optimal” because it was selected to work under some conditions and it is selected as long as the function is performed well enough. Following Jacob’s [246] metaphor, evolution is not an engineer that has access to all the materials needed to design a machine but resembles more the role of a tinkerer who does not know exactly what she/he is going to produce but uses everything at its disposal to produce some kind of workable object. An engineer depends on having the right raw materials, whereas a tinkerer manages it with odds and ends that can be used in a number of different ways. Thus, evolution most often creates novelty from pre-existing features, has no defined objective, and, thus, no aim for perfection—it has no aim at all; it yields workable solution (see also [246]). An example of molecular tinkering is the enzyme RuBisCO, which might have arisen in a non-autotrophic context (potentially for nucleotide assimilation) and was only later co-opted for fixing carbon in the Calvin-Benson cycle [247]. The study of metabolism evolution not only informs us on

adaptations in different lineages, but also on the historical constraints that shaped the evolution of pathways [18].

8.2. On by-products

Several compounds (especially molecules with signaling function) of the specialized metabolism might have once evolved from by-products since they are constantly formed under physiological conditions [248–251] and some, as of today, could still be considered by-products [204]. What do we mean by by-product formation? In general, we here consider four possible ways: A side product is formed in an enzymatic reaction as a “leftover” (Fig. 5, scenario 1); a thermodynamically and/or kinetically less favorable product is formed by an enzymatic reaction (Fig. 5, scenario 2); enzymatically formed products react spontaneously to more thermodynamically stable forms (Fig. 5, scenario 3); and/or environmental factors induce spontaneous degradation of biomolecules (Fig. 5, scenario 4; e.g. ROS-cleavage). It is conceivable that the latter two scenarios occur more profoundly in terrestrial habitats owing to the more intense thermodynamic and quantum-mechanic forces (as mentioned in the introduction).

In the end we cannot tell *why* some of these by-products have become elaborate or well-conserved signals, but we can raise the question: “What makes a by-product likely to become a specialized metabolic signal?” Here, we can rationalize some aspects on the basis of the scenarios mentioned (Fig. 5): if a by-product is formed in an enzymatic reaction every time or (the opposite) only under well-defined, but unusual conditions the foundation is laid that such by-product may become a specialized metabolic signal. The same is true if a spontaneously formed end-product is more stable under physiological conditions than the enzymatically formed product itself. Also, if an environmental factor always leads to the same by-product in a reliable manner (e.g. β -CC formed by ROS-cleavage of β -Carotene in a concentration-dependent manner) it might become part of the specialized metabolic chassis. As illustrated for apocarotenoids (see section above), their role in signaling is derived from the process they emerge from; that is, they are formed under oxidative stress and became over time a signal for the plant to induce a stress response towards oxidative stress. If (i) the recruitment of enzymes stabilizing the process occurred and (ii) a signaling protein

cascade exists/can be recruited, natural selection can act on both the signal and the output.

These processes continue to act on both a conserved and divergent pool of compounds; thus, the same compounds, related compounds from a similar source, but also those compounds sourced from distal/unrelated pathways might become major players in the molecular biology of an organism. All of these cases will have played out in the evolution of any extinct or extant lineage alike—and thus also since land plants and streptophyte algae diverged. The coming years will tell which scenarios have dominantly shaped which group of specialized metabolites—phytohormones, protectants, and more.

9. Concluding remarks: embracing the phylodiversity of specialized metabolism

Understanding the biochemical capacity of the earliest land plants necessitates investigations of streptophyte algae—which include the closest algal relatives of land plants; combined with phylodiverse data on land plants, robust inferences can be made. Nowadays, high quality genomes of representative members of all streptophyte algal classes—except Coleochaetophyceae—are available (for an overview, see [113,252]). Several complementation essays with streptophyte algal genes in genetically accessible land plants have been performed. This primarily allows for the elucidation of the conservation of function of several proteins, enzymes and specialized metabolites. Integrated into an evolutionary framework, the elucidation of ancestral functions can thus be performed. In order to move towards understanding the specialized metabolism of streptophyte algae—i.e., away from only understanding conservation and towards illuminating divergence—similar work in the algae themselves, is needed.

In recent time, some first transient and even stable transformation protocol for streptophyte algae have been published [117,253–257]. Studying biochemical responses across streptophyte diversity will provide insights into so far unknown regulatory processes. This way the evolutionary degrees of freedom for crucial metabolic inventions for greening planet Earth might become visible.

Our current view of plant specialized metabolism is biased—not least owing to the prominence of a few model systems in plant research. Streptophyte algae metabolomes illustrate this in an impressive manner. Several metabolites that were thought to be specific to embryophytes have been detected in streptophyte algae—even though orthologous sequences for crucial enzymes of the metabolism known from land plants (and most of the times from *A. thaliana*) are missing (e.g. phenylpropanoid-derived compounds, see [51]). A common inference is that the “complete” version of such pathways appeared only in embryophytes or even in angiosperms [258]. Examples of such pathways are not only the phenylpropanoid pathway, but also the ABA signaling pathway, the glucosinolate pathway, auxin signaling and transport and jasmonic acid [18,39,171,172,224,230–232,259]. This pattern has sometimes been proposed as evidence for the stepwise assembly of pathways and the complexity of specialized metabolism has even been proposed to correlate with increased structural complexity in plants [17]. However, looking for the presence of enzymes and compounds known from Arabidopsis across the diversity of plants and algae, although highly informative, has the peril of providing a biased view of evolution as proceeding along the phylogenetic backbone from algae to angiosperms rather than as a branching process. If we do not look for it, we cannot find it. This is true for metabolic and enzymatic diversity alike. Oftentimes such “incomplete pathways” are completed by different reactions and enzymes [18]. For example, the tricarboxylic acid cycle was thought to be incomplete in some cyanobacteria that lack the enzyme to convert 2-oxoglutarate into succinate, but this reaction was later shown to be catalyzed by two other enzymes (2-oxoglutarate decarboxylase and succinic demialdehyde dehydrogenase [260]). Some other pathways show the opposite pattern: the biosynthesis of diacylglyceryl-N,N,N-trimethylhomoserine (DGTS) is lost in

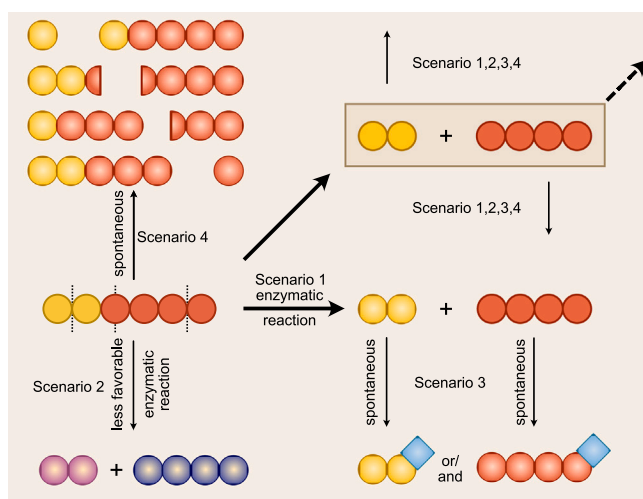


Fig. 5. Scenarios for the evolutionary origin of signals born out of by-products. By-products are gradient-colored. The main enzymatic reaction is depicted by a thicker reaction arrow. Scenario 1: A side product is formed in an enzymatic reaction as a “leftover”. Scenario 2: a thermodynamically and/or kinetically less favorable product is formed by an enzymatic reaction. Scenario 3: enzymatically formed products react spontaneously to more thermodynamically stable forms. Scenario 4: environmental factors induce spontaneous degradation of biomolecules (dashed lines; e.g., ROS-cleavage).

streptophytes and that of selenocysteine in embryophytes [258]. In cases where only a few enzymes are found in non-model organisms, gene co-expression networks can reveal the potential involvement of new enzymes in pathways, helping to build metabolic and signaling pathways in non-model organisms [261,262]. As a final word it is prudent to note that also cellular and sub-cellular structures for the accumulation of secondary metabolites can be found across hundreds of millions of years of evolution [263–265]—and we can only imagine what might be found in specific structures such as the highly resilient akinetes of filamentous Zygnematophyceae; although these are specific to certain lineages, their specialized metabolism might tell a tale of both co-option of conserved metabolic routes as well as unique metabolomes. Overall, a phylo-diverse perspective on metabolism will not only reflect the different biology of species—it will be highly informative on the effect of historical contingency in shaping the evolution of core and specialized metabolism.

Until now, systematic investigations into the chemodiversity of streptophyte algae are rare. Since specialized metabolites often appear to be lineage- or even species-specific, consecutive studies are needed to paint a complete picture of streptophyte biology. That is not to say that no generalistic conclusions can be drawn from species-specific metabolomes: several mechanisms and especially specialized metabolisms that define embryophyte biology already evolved *before* the conquest of land; the closest algal relatives of land plants do *not* comprise a “primitive” form *but* evolved their own (sometimes even more) elaborate sensing and mitigation strategies for environmental fluctuations. Now that the genomic data have set the stage, the time for the exploration of the specialized metabolome of the closest algal relatives of land plants has come.

Funding

J.d.V. thanks the European Research Council for funding under the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. 852725; ERC-StG “TerreStriAL”). J.d.V. is grateful for support through the German Research Foundation (DFG) within the framework of the Priority Program “MAdLand – Molecular Adaptation to Land: Plant Evolution to Change” (SPP 2237; VR 132/4–1), in which T.P.R. is a PhD student and A.D., A.D.A., J.M.R.F.-J, and I. I. partake as associate members.

Declaration of Competing Interest

We declare no competing interests.

Acknowledgments

T.P.R. and J.M.R.F.-J. gratefully acknowledge support by the Ph.D. program “Microbiology and Biochemistry” within the framework of the “Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences” (GGNB) at the University of Göttingen; A.D. and A.D.A. are grateful for being supported through the International Max Planck Research School (IMPRS) for Genome Science. We are grateful for pre-publication access to genome data of *Sphagnum fallax* (v1.1, DOE-JGI, <http://phytozome.jgi.doe.gov/>).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.semcd.2022.03.004](https://doi.org/10.1016/j.semcd.2022.03.004).

References

- [1] Y.M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth, *Proc. Natl. Acad. Sci. USA* 115 (2018) 6506–6511, <https://doi.org/10.1073/pnas.1711842115>.

- [2] L. Hoffmann, Algae of terrestrial habitats, *Bot. Rev.* 55 (1989) 77–105, <https://doi.org/10.1007/BF02858529>.
- [3] L.A. Lewis, P.O. Lewis, Unearthing the molecular phylo-diversity of desert soil green algae (Chlorophyta), *Syst. Biol.* 54 (2005) 936–947, <https://doi.org/10.1080/10635150500354852>.
- [4] A. Holzinger, U. Karsten, Desiccation stress and tolerance in green algae: Consequences for ultrastructure, physiological, and molecular mechanisms, *Front Plant Sci.* 4 (2013) 327, <https://doi.org/10.3389/fpls.2013.00327>.
- [5] J.L. Morris, M.N. Puttick, J.W. Clark, D. Edwards, P. Kenrick, S. Pressel, C. H. Wellman, Z. Yang, H. Schneider, P.C.J. Donoghue, The timescale of early land plant evolution, *Proc. Natl. Acad. Sci. U. S. A.* 115 (2018) E2274–E2283, <https://doi.org/10.1073/pnas.1719588115>.
- [6] C.F. Delwiche, E.D. Cooper, The evolutionary origin of a terrestrial flora, *Curr. Biol.* 25 (2015) R899–R910, <https://doi.org/10.1016/j.cub.2015.08.029>.
- [7] J. de Vries, J.M. Archibald, Plant evolution: landmarks on the path to terrestrial life, *N. Phytol.* 217 (2018) 1428–1434, <https://doi.org/10.1111/nph.14975>.
- [8] J.M.R. Fürst-Jansen, S. de Vries, J. de Vries, Evo-physio: on stress responses and the earliest land plants, *J. Exp. Bot.* 71 (2020) 3254–3269, <https://doi.org/10.1093/jxb/eraa007>.
- [9] S. de Vries, J. de Vries, J.K. von Dahlen, S.B. Gould, J.M. Archibald, L.E. Rose, C. H. Slamovits, On plant defense signaling networks and early land plant evolution, *Commun. Integr. Biol.* 11 (2018) 1–14, <https://doi.org/10.1080/19420889.2018.1486168>.
- [10] G.-Z. Han, Origin and evolution of the plant immune system, *N. Phytol.* 222 (2019) 70–83, <https://doi.org/10.1111/nph.15596>.
- [11] P.-M. Delaux, S. Schornack, Plant evolution driven by interactions with symbiotic and pathogenic microbes, *Science* 371 (2021) eaba6605, <https://doi.org/10.1126/science.aba6605>.
- [12] N.A. Bock, S. Charvet, J. Burns, Y. Gyaltsen, A. Rozenberg, S. Duhamel, E. Kim, Experimental identification and in silico prediction of bacterivory in green algae, *ISME J.* 15 (2021) 1987–2000, <https://doi.org/10.1038/s41396-021-00899-w>.
- [13] M. Abdelrhman, Modeling water clarity and light quality in oceans, *J. Mar. Sci. Eng. A* (2016) 80, <https://doi.org/10.3390/jmse4040080>.
- [14] H.C. Resemann, C. Herrfurth, K. Feussner, E. Hornung, A.K. Ostendorf, J. Gömann, J. Mittag, N. van Gessel, J. de Vries, J. Ludwig-Müller, J. Markham, R. Reski, I. Feussner, Convergence of sphingolipid desaturation across over 500 million years of plant evolution, *Nat. Plants* 7 (2021) 219–232, <https://doi.org/10.1038/s41477-020-00844-3>.
- [15] L.A. Lewis, R.M. McCourt, Green algae and the origin of land plants, *Am. J. Bot.* 91 (2004) 1535–1556, <https://doi.org/10.3732/ajb.91.10.1535>.
- [16] H. Ashihara, H. Sano, A. Crozier, Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering, *Phytochemistry* 69 (2008) 841–856, <https://doi.org/10.1016/j.phytochem.2007.10.029>.
- [17] J.-K. Weng, The evolutionary paths towards complexity: a metabolic perspective, *N. Phytol.* 201 (2014) 1141–1149, <https://doi.org/10.1111/nph.12416>.
- [18] H.A. Maeda, A.R. Fernie, Evolutionary history of plant metabolism, *Annu. Rev. Plant Biol.* 72 (2021) 185–216, <https://doi.org/10.1146/annurev-arplant-080620-031054>.
- [19] J.E. Baldwin, H. Krebs, The evolution of metabolic cycles, *Nature* 291 (1981) 381–382, <https://doi.org/10.1038/291381a0>.
- [20] R. Sugiyama, R. Li, A. Kuwahara, R. Nakabayashi, N. Sotta, T. Mori, T. Ito, N. Ohkama-Ohtsu, T. Fujiwara, K. Saito, R.T. Nakano, P. Bednarek, M.Y. Hirai, Retrograde sulfur flow from glucosinolates to cysteine in *Arabidopsis thaliana*, *Proc. Natl. Acad. Sci. USA* 118 (2021), <https://doi.org/10.1073/pnas.2017890118>.
- [21] P.-M. Delaux, G.V. Radhakrishnan, D. Jayaraman, J. Cheema, M. Malbreil, J. D. Volkering, H. Sekimoto, T. Nishiyama, M. Melkonian, L. Pokorny, C. J. Rothfels, H.W. Sederoff, D.W. Stevenson, B. Surek, Y. Zhang, M.R. Sussman, C. Dunand, R.J. Morris, C. Roux, G.K.-S. Wong, G.E.D. Oldroyd, J.-M. Ané, Algal ancestor of land plants was preadapted for symbiosis, *Proc. Natl. Acad. Sci. USA* 112 (2015) 13390–13395, <https://doi.org/10.1073/pnas.1515426112>.
- [22] J.J. Knack, L.W. Wilcox, P.M. Delaux, J.M. Ané, M.J. Piotrowski, M.E. Cook, J. M. Graham, L.E. Graham, Microbiomes of streptophyte algae and bryophytes suggest that a functional suite of microbiota fostered plant colonization of land, *Int. J. Plant Sci.* 176 (2015) 405–420, <https://doi.org/10.1086/681161>.
- [23] M.L. Berbee, T.Y. James, C. Strullu-Derrien, Early diverging fungi: diversity and impact at the dawn of terrestrial life, *Annu. Rev. Microbiol.* 71 (2017) 41–60, <https://doi.org/10.1146/annurev-micro-030117-020324>.
- [24] M.L. Berbee, C. Strullu-Derrien, P.-M. Delaux, P.K. Strother, P. Kenrick, M.-A. Selosse, J.W. Taylor, Genomic and fossil windows into the secret lives of the most ancient fungi, *Nat. Rev. Microbiol.* 18 (2020) 717–730, <https://doi.org/10.1038/s41579-020-0426-8>.
- [25] Y. Gao, W. Wang, T. Zhang, Z. Gong, H. Zhao, G.-Z. Han, Out of water: the origin and early diversification of plant R-genes, *Plant Physiol.* 177 (2018) 82–89, <https://doi.org/10.1104/pp.18.00185>.
- [26] A. Dievart, C. Götting, C. Périn, V. Ranwez, N. Chantret, Origin and diversity of plant receptor-like kinases, *Annu. Rev. Plant Biol.* 71 (2020) 131–156, <https://doi.org/10.1146/annurev-arplant-073019-025927>.
- [27] Z. Gong, G. Han, Flourishing in water: the early evolution and diversification of plant receptor-like kinases, *Plant J.* 106 (2021) 174–184, <https://doi.org/10.1111/tj.15157>.
- [28] Z.A. Popper, G. Michel, C. Hervé, D.S. Domozych, W.G.T. Willats, M.G. Tuohy, B. Kloareg, D.B. Stengel, Evolution and Diversity of Plant Cell Walls: From Algae to Flowering Plants, *Annu. Rev. Plant Biol.* 62 (2011) 567–590, <https://doi.org/10.1146/annurev-arplant-042110-103809>.

- [29] S.A. Rensing, Great moments in evolution: the conquest of land by plants, *Curr. Opin. Plant Biol.* 42 (2018) 49–54, <https://doi.org/10.1016/j.pbi.2018.02.006>.
- [30] O. Kherssonsky, D.S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective, *Annu Rev. Biochem.* 79 (2010) 471–505, <https://doi.org/10.1146/annurev-biochem-030409-143718>.
- [31] R. D'Ari, J. Casadesús, Underground metabolism, *Bioessays* 20 (1998) 181–186, [https://doi.org/10.1002/\(SICI\)1521-1878\(199802\)20:2<181::AID-BIES10>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<181::AID-BIES10>3.0.CO;2-0).
- [32] N. Kashtan, E. Noor, U. Alon, Varying environments can speed up evolution, *Proc. Natl. Acad. Sci. USA.* 104 (2007) 13711–13716, <https://doi.org/10.1073/pnas.0611630104>.
- [33] L. Jones, A.R. Ennos, S.R. Turner, Cloning and characterization of *irregular xylem4* (*irx4*): a severely lignin-deficient mutant of *Arabidopsis*, *Plant J.* 26 (2001) 205–216, <https://doi.org/10.1046/j.1365-313x.2001.01021.x>.
- [34] J. Thévenin, B. Pollet, B. Letarnec, L. Saulnier, L. Gissot, A. Maia-Grandard, C. Lapiere, L. Jouanin, The Simultaneous Repression of CCR and CAD, Two Enzymes of the Lignin Biosynthetic Pathway, Results in Sterility and Dwarfism in *Arabidopsis thaliana*, *Mol. Plant* 4 (2011) 70–82, <https://doi.org/10.1093/mp/ssq045>.
- [35] M. Lee, H.S. Jeon, S.H. Kim, J.H. Chung, D. Roppolo, H. Lee, H.J. Cho, Y. Tobimatsu, J. Ralph, O.K. Park, Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants, *EMBO J.* 38 (2019), 101948, <https://doi.org/10.15252/embj.2019101948>.
- [36] R.A. Dixon, N.L. Paiva, Stress-induced phenylpropanoid metabolism, *Plant Cell* 7 (1995) 1085–1097, <https://doi.org/10.1105/tpc.7.7.1085>.
- [37] D. Nelson, D. Werck-Reichhart, A P450-centric view of plant evolution: P450-centric evolution, *Plant J.* 66 (2011) 194–211, <https://doi.org/10.1111/j.1365-313x.2011.04529.x>.
- [38] A.V. Alber, H. Renault, A. Basilio-Lopes, J. Bassard, Z. Liu, P. Ullmann, A. Lesot, F. Bihel, M. Schmitt, D. Werck-Reichhart, J. Ehling, Evolution of coumaroyl conjugate 3-hydroxylases in land plants: lignin biosynthesis and defense, *Plant J.* (2019) 14373, <https://doi.org/10.1111/tpj.14373>.
- [39] S. de Vries, J.M.R. Fürst-Jansen, I. Irisarri, A. Dhabalia Ashok, T. Ischebeck, K. Feussner, I.N. Abreu, M. Petersen, J. de Vries, The evolution of the phenylpropanoid pathway entailed pronounced radiation and divergence of enzyme families, *Plant J.* 107 (2021) 975–1002, <https://doi.org/10.1111/tpj.15387>.
- [40] G. Emiliani, M. Fondi, R. Fani, S. Gribaldo, A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land, *Biol. Direct* 4 (2009) 7, <https://doi.org/10.1186/1745-6150-4-7>.
- [41] L. Labeeuw, P.T. Martone, Y. Boucher, R.J. Case, Ancient origin of the biosynthesis of lignin precursors, *Biol. Direct* 10 (2015) 23, <https://doi.org/10.1186/s13062-015-0052-y>.
- [42] J. de Vries, S. de Vries, C.H. Slamovits, L.E. Rose, J.M. Archibald, How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant Cell Physiol.* 58 (2017) 934–945, <https://doi.org/10.1093/pcp/pcx037>.
- [43] C.F. Delwiche, L.E. Graham, N. Thomson, Lignin-like compounds and sporopollenin in Coleochaete, an algal model for land plant ancestry, *Science* 245 (1989) 399–401, <https://doi.org/10.1126/science.245.4916.399>.
- [44] S.B. Krokene, L.E. Graham, M.E. Cook, Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes, *Am. J. Bot.* 83 (1996) 1241–1254, <https://doi.org/10.1002/j.1537-2197.1996.tb13908.x>.
- [45] R. Ligrone, A. Carafa, J.G. Duckett, K.S. Renzaglia, K. Ruel, Immunocytochemical detection of lignin-related epitopes in cell walls in bryophytes and the charalean alga *Nitzella*, *Plant Syst. Evol.* 270 (2008) 257–272, <https://doi.org/10.1007/s00606-007-0617-z>.
- [46] I. Sørensen, F.A. Pettolino, A. Bacic, J. Ralph, F. Lu, M.A. O'Neill, Z. Fei, J.K. Rose, D.S. Domozych, W.G.T. Willats, The charophycean green algae provide insights into the early origins of plant cell walls: cell-wall evolution and the charophycean green algae, *Plant J.* 68 (2011) 201–211, <https://doi.org/10.1111/j.1365-313x.2011.04686.x>.
- [47] K. Goiris, K. Muylaert, S. Voorspoels, B. Noten, D. De Paep, G.J.E. Baart, L. De Cooman, Detection of flavonoids in microalgae from different evolutionary lineages, *J. Phycol.* 50 (2014) 483–492, <https://doi.org/10.1111/jpy.12180>.
- [48] K. Hori, F. Maruyama, T. Fujisawa, T. Togashi, N. Yamamoto, M. Seo, S. Sato, T. Yamada, H. Mori, N. Tajima, T. Moriyama, M. Ikeuchi, M. Watanabe, H. Wada, K. Kobayashi, M. Saito, T. Masuda, Y. Sasaki-Sekimoto, K. Mashiguchi, K. Awai, M. Shimojima, S. Masuda, M. Iwai, T. Nobusawa, T. Narise, S. Kondo, H. Saito, R. Sato, M. Murakawa, Y. Ihara, Y. Oshima-Yamada, K. Ohtaka, M. Satoh, K. Sonobe, M. Ishii, R. Ohtani, M. Kanamori-Sato, R. Honoki, D. Miyazaki, H. Mochizuki, J. Umetsu, K. Higashi, D. Shibata, Y. Kamiya, N. Sato, Y. Nakamura, S. Tabata, S. Ida, K. Kurokawa, H. Ohta, *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation, *Nat. Commun.* 5 (2014) 3978, <https://doi.org/10.1038/ncomms4978>.
- [49] J. de Vries, S. Vries, B.A. Curtis, H. Zhou, S. Penny, K. Feussner, D.M. Pinto, M. Steinert, A.M. Cohen, K. Schwartzberg, J.M. Archibald, Heat stress response in the closest algal relatives of land plants reveals conserved stress signaling circuits, *Plant J.* 103 (2020) 1025–1048, <https://doi.org/10.1111/tpj.14782>.
- [50] One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants, *Nature* 574 (2019) 679–685, <https://doi.org/10.1038/s41586-019-1693-2>.
- [51] C. Jiao, I. Sørensen, X. Sun, H. Sun, H. Behar, S. Alseekh, G. Philippe, K. Palacio Lopez, L. Sun, R. Reed, S. Jeon, R. Kiyonami, S. Zhang, A.R. Fernie, H. Brumer, D. S. Domozych, Z. Fei, J.K.C. Rose, The *Penium margaritaceum* genome: hallmarks of the origins of land plants, *Cell* 181 (2020) 1097–1111.e12, <https://doi.org/10.1016/j.cell.2020.04.019>.
- [52] J. Barros, J.C. Serrani-Yarce, F. Chen, D. Baxter, B.J. Venables, R.A. Dixon, Role of bifunctional ammonia-lyase in grass cell wall biosynthesis, *Nat. Plants* 2 (2016) 16050, <https://doi.org/10.1038/nplants.2016.50>.
- [53] J. Barros, R.A. Dixon, Plant Phenylalanine/Tyrosine Ammonia-lyases, *Trends Plant Sci.* 25 (2020) 66–79, <https://doi.org/10.1016/j.tplants.2019.09.011>.
- [54] J.-K. Weng, X. Li, J. Stout, C. Chapple, Independent origins of syringyl lignin in vascular plants, *Proc. Natl. Acad. Sci. USA.* 105 (2008) 7887–7892, <https://doi.org/10.1073/pnas.0801696105>.
- [55] J.-K. Weng, C. Chapple, The origin and evolution of lignin biosynthesis, *N. Phytol.* 187 (2010) 273–285, <https://doi.org/10.1111/j.1469-8137.2010.03327.x>.
- [56] J.-K. Weng, T. Akiyama, J. Ralph, C. Chapple, Independent recruitment of an O-methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*, *Plant Cell* 23 (2011) 2708–2724, <https://doi.org/10.1105/tpc.110.081547>.
- [57] B.T. Piatkowski, K. Imwattana, E.A. Tripp, D.J. Weston, A. Healey, J. Schmutz, A. J. Shaw, Phylogenomics reveals convergent evolution of red-violet coloration in land plants and the origins of the anthocyanin biosynthetic pathway, *Mol. Phylogenet. Evol.* 151 (2020), 106904, <https://doi.org/10.1016/j.ympev.2020.106904>.
- [58] S. Aigner, D. Remias, U. Karsten, A. Holzinger, Unusual phenolic compounds contribute to ecophysiological performance in the purple-colored green alga *Zygonium ericetorum* (Zygnematophyceae, Streptophyta) from a high-alpine habitat, *J. Phycol.* 49 (2013) 648–660, <https://doi.org/10.1111/jpy.12075>.
- [59] M. Pichrtová, D. Remias, L.A. Lewis, A. Holzinger, Changes in phenolic compounds and cellular ultrastructure of arctic and antarctic strains of *Zygnema* (Zygnematophyceae, Streptophyta) after exposure to experimentally enhanced UV to PAR ratio, *Micro Ecol.* 65 (2013) 68–83, <https://doi.org/10.1007/s00248-012-0096-9>.
- [60] A. Holzinger, A. Albert, S. Aigner, J. Uhl, P. Schmitt-Kopplin, K. Trumhová, M. Pichrtová, Arctic, Antarctic, and temperate green algae *Zygnema* spp. under UV-B stress: vegetative cells perform better than pre-akinetes, *Protoplasma* 255 (2018) 1239–1252, <https://doi.org/10.1007/s00709-018-1225-1>.
- [61] C.M. Furlan, L. Motta, D. Santos, Tannins: what do they represent in plant life?, in: *Tannins: Types, Foods Containing, and Nutrition*. (2011) pp. 251–263.
- [62] A. Busch, S. Hess, Sunscreen mucilage: a photoprotective adaptation found in terrestrial green algae (Zygnematophyceae), *Eur. J. Phycol.* (2021) 1–18, <https://doi.org/10.1080/09670262.2021.1898677>.
- [63] E. Litchman, P.J. Neale, A.T. Banaszak, Increased sensitivity to ultraviolet radiation in nitrogen-limited dinoflagellates: Photoprotection and repair, *Limnol. Oceanogr.* 47 (2002) 86–94, <https://doi.org/10.4319/lo.2002.47.1.0086>.
- [64] U. Karsten, A. Holzinger, Green algae in alpine biological soil crust communities: acclimation strategies against ultraviolet radiation and dehydration, *Biodivers. Conserv.* 23 (2014) 1845–1858, <https://doi.org/10.1007/s10531-014-0653-2>.
- [65] A. Hartmann, K. Glaser, A. Holzinger, M. Ganzera, U. Karsten, Klebsormidin A and b, two new uv-sunscreen compounds in green microalgal *Interfilum* and *Klebsormidium* species (streptophyta) from terrestrial habitats, *Front. Microbiol.* 11 (2020) 499, <https://doi.org/10.3389/fmicb.2020.00499>.
- [66] U. Karsten, K. Herburger, A. Holzinger, Dehydration, temperature, and light tolerance in members of the aeroterrestrial green algal genus *Interfilum* (Streptophyta) from biogeographically different temperate soils, *J. Phycol.* 50 (2014) 804–816, <https://doi.org/10.1111/jpy.12210>.
- [67] H. Renaut, A. Alber, N.A. Horst, A. Basilio Lopes, E.A. Fich, L. Kriegshauser, G. Wiedemann, P. Ullmann, L. Herrgott, M. Erhardt, E. Pineau, J. Ehling, M. Schmitt, J.K.C. Rose, R. Reski, D. Werck-Reichhart, A phenol-enriched cuticle is ancestral to lignin evolution in land plants, *Nat. Commun.* 8 (2017) 14713, <https://doi.org/10.1038/ncomms14713>.
- [68] P. de Vries, J. Simons, A.P. van Beem, Sporopollenin in the spore wall of *Spirogyra* (Zygnemataceae, Chlorophyceae), *Acta Bot. Neerl.* 32 (1983) 25–28, <https://doi.org/10.1111/j.1438-8677.1983.tb01674.x>.
- [69] S. Blackmore, S.H. Barnes, Pollen wall morphogenesis in *Tragopogon porrifolius* L. (Compositae: Lactuceae) and its taxonomic significance, *Rev. Palaeobot. Palynol.* 52 (1987) 233–246, [https://doi.org/10.1016/0034-6667\(87\)90056-X](https://doi.org/10.1016/0034-6667(87)90056-X).
- [70] H. Berland, N.W. Albert, A. Stavland, M. Jordheim, T.K. McGhie, Y. Zhou, H. Zhang, S.C. Derolis, K.E. Schwinn, B.R. Jordan, K.M. Davies, Ø.M. Andersen, Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants, *Proc. Natl. Acad. Sci. USA.* 116 (2019) 20232–20239, <https://doi.org/10.1073/pnas.1912741116>.
- [71] H. Renault, D. Werck-Reichhart, J.-K. Weng, Harnessing lignin evolution for biotechnological applications, *Curr. Opin. Biotechnol.* 56 (2019) 105–111, <https://doi.org/10.1016/j.copbio.2018.10.011>.
- [72] P. Ding, Y. Ding, Stories of Salicylic Acid: A Plant Defense Hormone, *Trends Plant Sci.* 25 (2020) 549–565, <https://doi.org/10.1016/j.tplants.2020.01.004>.
- [73] P. Meuwly, W. Molders, A. Buchala, J.P. Métraux, Local and systemic biosynthesis of salicylic acid in infected cucumber plants, *Plant Physiol.* 109 (1995) 1107–1114, <https://doi.org/10.1104/pp.109.3.1107>.
- [74] J.A. Pallas, N.L. Paiva, C. Lamb, R.A. Dixon, Tobacco plants epigenetically suppressed in phenylalanine ammonia-lyase expression do not develop systemic acquired resistance in response to infection by tobacco mosaic virus, *Plant J.* 10 (1996) 281–293, <https://doi.org/10.1046/j.1365-313x.1996.10020281.x>.
- [75] J.-L. Coquoz, A. Buchala, J.-P. Métraux, The biosynthesis of salicylic acid in potato, *Plants, Plant Physiol.* 117 (1998) 1095–1101, <https://doi.org/10.1104/pp.117.3.1095>.
- [76] C. Garcion, A. Lohmann, E. Lamodièrre, J. Catinot, A. Buchala, P. Doermann, J.-P. Métraux, Characterization and biological function of the *ISOCHORISMATE*

- SYNTHASE2 gene of *Arabidopsis*, *Plant Physiol.* 147 (2008) 1279–1287, <https://doi.org/10.1104/pp.108.119420>.
- [77] Z. Chen, Z. Zheng, J. Huang, Z. Lai, B. Fan, Biosynthesis of salicylic acid in plants, *Plant Signal. Behav.* 4 (2009) 493–496, <https://doi.org/10.4161/psb.4.6.8392>.
- [78] M.C. Wildermuth, J. Dewdney, G. Wu, F.M. Ausubel, Isochorismate synthase is required to synthesize salicylic acid for plant defence, *Nature* 414 (2001) 562–565, <https://doi.org/10.1038/35107108>.
- [79] M. Serrano, B. Wang, B. Aryal, C. Garcion, E. Abou-Mansour, S. Heck, M. Geisler, F. Mauch, C. Nawrath, J.-P. Métraux, Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5, *Plant Physiol.* 162 (2013) 1815–1821, <https://doi.org/10.1104/pp.113.218156>.
- [80] D. Rekhter, D. Lüdke, Y. Ding, K. Feussner, K. Zienkiewicz, V. Lipka, M. Wiermer, Y. Zhang, I. Feussner, Isochorismate-derived biosynthesis of the plant stress hormone salicylic acid, *Science* 365 (2019) 498–502, <https://doi.org/10.1126/science.aaw1720>.
- [81] M.P. Torrens-Spence, A. Bobokalonova, V. Carballo, C.M. Glinkerman, T. Pluskal, A. Shen, J.-K. Weng, PBS3 and EPS1 complete salicylic acid biosynthesis from isochorismate in *Arabidopsis*, *Mol. Plant* 12 (2019) 1577–1586, <https://doi.org/10.1016/j.molp.2019.11.005>.
- [82] M.P. Torrens-Spence, T. Li, Z. Wang, C.M. Glinkerman, J.O. Matos, Y. Wang, J.-K. Weng, Mechanistic basis for the emergence of EPS1 as a catalyst in plant salicylic acid biosynthesis, *bioRxiv* (2021), <https://doi.org/10.1101/2021.08.21.457228>.
- [83] M.B. Shine, J. Yang, M. El-Habbak, P. Nagyabhyru, D. Fu, D. Navarre, S. Ghabrial, P. Kachroo, A. Kachroo, Cooperative functioning between phenylalanine ammonia lyase and isochorismate synthase activities contributes to salicylic acid biosynthesis in soybean, *N. Phytol.* 212 (2016) 627–636, <https://doi.org/10.1111/nph.14078>.
- [84] J. Zeier, Metabolic regulation of systemic acquired resistance, *Curr. Opin. Plant Biol.* 62 (2021), 102050, <https://doi.org/10.1016/j.cpb.2021.102050>.
- [85] J. Catinot, A. Buchala, E. Abou-Mansour, J.-P. Métraux, Salicylic acid production in response to biotic and abiotic stress depends on isochorismate in *Nicotiana benthamiana*, *FEBS Lett.* 582 (2008) 473–478, <https://doi.org/10.1016/j.febslet.2007.12.039>.
- [86] D.S. Kim, B.K. Hwang, An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens, *J. Exp. Bot.* 65 (2014) 2295–2306, <https://doi.org/10.1093/jxb/eru109>.
- [87] J.R. Widhalm, N. Dudareva, A familiar ring to it: biosynthesis of plant benzoic acids, *Mol. Plant* 8 (2015) 83–97, <https://doi.org/10.1016/j.molp.2014.12.001>.
- [88] M. Ibdah, Y.-T. Chen, C.G. Wilkerson, E. Pichersky, An aldehyde oxidase in developing seeds of *Arabidopsis* converts benzaldehyde to benzoic acid, *Plant Physiol.* 150 (2009) 416–423, <https://doi.org/10.1104/pp.109.135848>.
- [89] M.C. Long, D.A. Nagegowda, Y. Kaminaga, K.K. Ho, C.M. Kish, J. Schnepf, D. Sherman, H. Weiner, D. Rhodes, N. Dudareva, Involvement of snapdragon benzaldehyde dehydrogenase in benzoic acid biosynthesis, *Plant J.* 59 (2009) 256–265, <https://doi.org/10.1111/j.1365-313X.2009.03864.x>.
- [90] T.A. Colquhoun, D.M. Marciniak, A.E. Wedde, J.Y. Kim, M.L. Schwieterman, L.A. Levin, A. Van Moerkercke, R.C. Schuurink, D.G. Clark, A peroxisomally localized acyl-activating enzyme is required for volatile benzenoid formation in a *Petunia*×*hybrida* cv. ‘Mitchell Diploid’ flower, *J. Exp. Bot.* 63 (2012) 4821–4833, <https://doi.org/10.1093/jxb/ers153>.
- [91] A. Klempien, Y. Kaminaga, A. Qualley, D.A. Nagegowda, J.R. Widhalm, I. Orlova, A.K. Shasany, G. Taguchi, C.M. Kish, B.R. Cooper, J.C. D’Auria, D. Rhodes, E. Pichersky, N. Dudareva, Contribution of CoA Ligases to Benzenoid Biosynthesis in *Petunia* Flowers, *Plant Cell* 24 (2012) 2015–2030, <https://doi.org/10.1105/tpc.112.097519>.
- [92] S. Lee, Y. Kaminaga, B. Cooper, E. Pichersky, N. Dudareva, C. Chapple, Benzylation and sinapoylation of glucosinolate R-groups in *Arabidopsis*, *Plant J.* 72 (2012) 411–422, <https://doi.org/10.1111/j.1365-313X.2012.05096.x>.
- [93] A. Van Moerkercke, I. Schaubinhold, E. Pichersky, M.A. Haring, R.C. Schuurink, A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production, *Plant J.* 60 (2009) 292–302, <https://doi.org/10.1111/j.1365-313X.2009.03953.x>.
- [94] A.V. Qualley, J.R. Widhalm, F. Adebessin, C.M. Kish, N. Dudareva, Completion of the core-oxidative pathway of benzoic acid biosynthesis in plants, *Proc. Natl. Acad. Sci. USA* 109 (2012) 16383–16388, <https://doi.org/10.1073/pnas.1211001109>.
- [95] J.D. Bussell, M. Reichelt, A.A.G. Wiszniewski, J. Gershenzon, S.M. Smith, Peroxisomal ATP-binding cassette transporter COMATOSE and the multifunctional protein ABNORMAL INFLORESCENCE MERISTEM are required for the production of benzoylated metabolites in *Arabidopsis* seeds, *Plant Physiol.* 164 (2014) 48–54, <https://doi.org/10.1104/pp.113.229807>.
- [96] F. Adebessin, J.R. Widhalm, J.H. Lynch, R.M. McCoy, N. Dudareva, A peroxisomal thioesterase plays auxiliary roles in plant β -oxidative benzoic acid metabolism, *Plant J.* 93 (2018) 905–916, <https://doi.org/10.1111/tip.13818>.
- [97] J.R. Widhalm, A.-L. Ducluzeau, N.E. Buller, C.G. Elowsky, L.J. Olsen, G.J. C. Basset, Phyloquinone (vitamin K1) biosynthesis in plants: two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa: Plant DHNA-CoA thioesterases, *Plant J.* 71 (2012) 205–215, <https://doi.org/10.1111/j.1365-313X.2012.04972.x>.
- [98] J. León, V. Shulaev, N. Yalpani, M.A. Lawton, I. Raskin, Benzoic acid 2-hydroxylase, a soluble oxygenase from tobacco, catalyzes salicylic acid biosynthesis, *Proc. Natl. Acad. Sci. USA.* 92 (1995) 10413–10417, <https://doi.org/10.1073/pnas.92.22.10413>.
- [99] S. de Vries, C. Herrfurth, F.-W. Li, I. Feussner, J. de Vries, An ancient route towards salicylic acid and its implications for the perpetual *Trichormus–Azolla* symbiosis, *bioRxiv Prepr.* (2021), <https://doi.org/10.1101/2021.03.12.435107>.
- [100] L. Závěská Drábková, P.I. Dobrev, V. Motyka, Phytohormone profiling across the bryophytes, *PLoS ONE* 10 (2015), e0125411, <https://doi.org/10.1371/journal.pone.0125411>.
- [101] C. Wang, Y. Liu, S.-S. Li, G.-Z. Han, Insights into the origin and evolution of the plant hormone signaling machinery, *Plant Physiol.* 167 (2015) 872–886, <https://doi.org/10.1104/pp.114.247403>.
- [102] O. De Clerck, S.-M. Kao, K.A. Bogaert, J. Blomme, F. Foflonker, M. Kwantes, E. Vancaester, L. Vanderstraeten, E. Aydogdu, J. Boesger, G. Califano, B. Charrier, R. Clewes, A. Del Cortona, S. D’Hondt, N. Fernandez-Pozo, C.M. Gachon, M. Hanikenne, L. Lattermann, F. Leliart, X. Liu, C.A. Maggs, Z.A. Popper, J. A. Raven, M. Van Bel, P.K.I. Wilhelmsson, D. Bhattacharya, J.C. Coates, S. A. Rensing, D. Van Der Straeten, A. Vardi, L. Sterck, K. Vandepoele, Y. Van de Peer, T. Wichard, J.H. Bothwell, Insights into the evolution of multicellularity from the sea lettuce genome, *Curr. Biol.* 28 (2018) 2921–2933.e5, <https://doi.org/10.1016/j.cub.2018.08.015>.
- [103] T. Nishiyama, H. Sakayama, J. de Vries, H. Buschmann, D. Saint-Marcoux, K. K. Ullrich, F.B. Haas, L. Vanderstraeten, D. Becker, D. Lang, S. Vosolsobé, S. Rombauts, P.K.I. Wilhelmsson, P. Janitza, R. Kern, A. Heyl, F. Rümpler, L.I.A. C. Villalobos, J.M. Clay, R. Skokan, A. Toyoda, Y. Suzuki, H. Kagoshima, E. Schijlen, N. Tajeshwar, B. Catarino, A.J. Hetherington, A. Saltykova, C. Bonnot, H. Breuning, A. Symeonidi, G.V. Radhakrishnan, F. Van Nieuwerburgh, D. Deforce, C. Chang, K.G. Karol, H. Hedrich, P. Ulvskov, G. Glöckner, C. F. Delwiche, J. Petrásék, Y. Van de Peer, J. Friml, M. Beilby, L. Dolan, Y. Kohara, S. Sugano, A. Fujiyama, P.-M. Delaux, M. Quint, G. Theißen, M. Hagemann, J. Harholt, C. Dunand, S. Zachgo, J. Langdale, F. Maumus, D. Van Der Straeten, S. B. Gould, S.A. Rensing, The *Chara* genome: secondary complexity and implications for plant terrestrialization, *Cell* 174 (2018) 448–464.e24, <https://doi.org/10.1016/j.cell.2018.06.033>.
- [104] F.-W. Li, T. Nishiyama, M. Waller, E. Frangedakis, J. Keller, Z. Li, N. Fernandez-Pozo, M.S. Barker, T. Bennett, M.A. Blázquez, S. Cheng, A.C. Cuming, J. de Vries, S. de Vries, P.-M. Delaux, I.S. Diop, C.J. Harrison, D. Hauser, J. Hernández-García, A. Kirbis, J.C. Meeks, I. Monte, S.K. Murre, A. Neubauer, D. Quandt, T. Robison, M. Shimamura, S.A. Rensing, J.C. Villarreal, D. Weijers, S. Wicke, G.K.-S. Wong, K. Sakakibara, P. Szövényi, *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts, *Nat. Plants* 6 (2020) 259–272, <https://doi.org/10.1038/s41477-020-0618-2>.
- [105] J. Gross, W.K. Cho, L. Lezhneva, J. Falk, K. Krupinska, K. Shinozaki, M. Seki, R. G. Herrmann, J. Meurer, A Plant locus essential for phyloquinone (vitamin K1) biosynthesis originated from a fusion of four eubacterial genes, *J. Biol. Chem.* 281 (2006) 17189–17196, <https://doi.org/10.1074/jbc.M601754200>.
- [106] A.J. Toribio, F. Suárez-Estrella, M.M. Jurado, M.J. López, J.A. López-González, J. Moreno, Prospection of cyanobacteria producing bioactive substances and their application as potential phyto-stimulating agents, *Biotechnol. Rep.* 26 (2020), e00449, <https://doi.org/10.1016/j.btre.2020.e00449>.
- [107] R.G. Ankenbauer, C.D. Cox, Isolation and characterization of *Pseudomonas aeruginosa* mutants requiring salicylic acid for pyochelin biosynthesis, *J. Bacteriol.* 170 (1988) 5364–5367, <https://doi.org/10.1128/jb.170.11.5364-5367.1988>.
- [108] C. Gaille, P. Kast, D. Haas, Salicylate biosynthesis in *Pseudomonas aeruginosa*, *J. Biol. Chem.* 277 (2002) 21768–21775, <https://doi.org/10.1074/jbc.M20410200>.
- [109] J. Zwahlen, S. Kolappan, R. Zhou, C. Kisker, P.J. Tonge, Structure and mechanism of MbtI, the salicylate synthase from *Mycobacterium tuberculosis*, *Biochemistry* 46 (2007) 954–964, <https://doi.org/10.1021/bi060852x>.
- [110] R. Daruwala, O. Kwon, R. Meganathan, M.E.S. Hudspeth, A new isochorismate synthase specifically involved in menaquinone (vitamin K2) biosynthesis encoded by the menF gene, *FEMS Microbiol. Lett.* 140 (1996) 159–163, <https://doi.org/10.1111/j.1574-6968.1996.tb08330.x>.
- [111] R. Meganathan, O. Kwon, Biosynthesis of Menaquinone (Vitamin K 2) and Ubiquinone (Coenzyme Q), *EcoSal* 3 (2009), <https://doi.org/10.1128/ecosalplus.3.6.3.3>.
- [112] E. Frangedakis, M. Waller, T. Nishiyama, H. Tsukaya, X. Xu, Y. Yue, M. Tjahjadi, A. Gunadi, J. Van Eck, F. Li, P. Szövényi, K. Sakakibara, An *Agrobacterium*-mediated stable transformation technique for the hornwort model *Anthoceros agrestis*, *New Phytol.* (2021) 17524, <https://doi.org/10.1111/nph.17524>.
- [113] P. Szövényi, A. Gunadi, F.-W. Li, Charting the genomic landscape of seed-free plants, *Nat. Plants* 7 (2021) 554–565, <https://doi.org/10.1038/s41477-021-00888-z>.
- [114] J.L. Bowman, T. Kohchi, K.T. Yamato, J. Jenkins, S. Shu, K. Ishizaki, S. Yamaoka, R. Nishihama, Y. Nakamura, F. Berger, C. Adam, S.S. Aki, F. Althoff, T. Araki, M. A. Arteaga-Vazquez, S. Balasubramanian, K. Barry, D. Bauer, C.R. Boehm, L. Briginshaw, J. Caballero-Pérez, B. Catarino, F. Chen, S. Chiyoda, M. Chovatia, K.M. Davies, M. Delmans, T. Demura, T. Dierschke, L. Dolan, A.E. Dorantes-Acosta, D.M. Eklund, S.N. Florent, E. Flores-Sandoval, A. Fujiyama, H. Fukuzawa, B. Galik, D. Grimaneli, J. Grimwood, U. Grossniklaus, T. Hamada, J. Haseloff, A. J. Hetherington, A. Higo, Y. Hirakawa, H.N. Hundley, Y. Ikeda, K. Inoue, S. Inoue, S. Ishida, Q. Jia, M. Kakita, T. Kanazawa, Y. Kawai, T. Kawashima, M. Kennedy, K. Kinoshita, T. Kinoshita, Y. Kohara, E. Koide, K. Komatsu, S. Kopschke, M. Kubo, J. Kyojuka, U. Lagercrantz, S.-S. Lin, E. Lindquist, A.M. Lipzen, C.-W. Lu, E. De Luna, R.A. Martienssen, N. Minamino, M. Mizutani, M. Mizutani, N. Mochizuki, I. Monte, R. Mosher, H. Nagasaki, H. Nakagami, S. Naramoto, K. Nishitani, M. Ohtani, T. Okamoto, M. Okumura, J. Phillips, B. Pollak, A. Reinders, M. Rövekamp, R. Sano, S. Sawa, M.W. Schmid, M. Shirakawa, R. Solano,

- A. Spunde, N. Suetsugu, S. Sugano, A. Sugiyama, R. Sun, Y. Suzuki, M. Takenaka, D. Takezawa, H. Tomogane, M. Tsuzuki, T. Ueda, M. Umeda, J.M. Ward, Y. Watanabe, K. Yazaki, R. Yokoyama, Y. Yoshitake, I. Yotsui, S. Zachgo, J. Schmutz, Insights into land plant evolution garnered from the *Marchantia polymorpha* genome, *Cell* 171 (2017) 287–299.e15, <https://doi.org/10.1016/j.cell.2017.09.030>.
- [115] S. Cheng, W. Xian, Y. Fu, B. Marin, J. Keller, T. Wu, W. Sun, X. Li, Y. Xu, Y. Zhang, S. Wittek, T. Reeder, G. Günther, A. Gontcharov, S. Wang, L. Li, X. Liu, J. Wang, H. Yang, X. Xu, P.-M. Delaux, B. Melkonian, G. Wong, M. Melkonian, Genomes of subaerial Zygnematophyceae provide insights into land plant evolution, *Cell* 179 (2019) 1057–1067, <https://doi.org/10.1016/j.cell.2019.10.019>.
- [116] S.A. Rensing, B. Goffinet, R. Meyberg, S.-Z. Wu, M. Bezanilla, The moss *Physcomitrium* (*Physcomitrella*) *patens*: a model organism for non-seed plants, *Plant Cell* 32 (2020) 1361–1376, <https://doi.org/10.1105/tpc.19.00828>.
- [117] J. Kawai, M. Kanazawa, R. Suzuki, N. Kikuchi, Y. Hayakawa, H. Sekimoto, Highly efficient transformation of the model zygnematophycean alga *Closterium peracerosum-strigosum-litorale* complex by square-pulse electroporation, *New Phytol.* (2021) 17763, <https://doi.org/10.1111/nph.17763>.
- [118] J.S. Thaler, P.T. Humphrey, N.K. Whiteman, Evolution of jasmonate and salicylate signal crosstalk, *Trends Plant Sci.* 17 (2012) 260–270, <https://doi.org/10.1016/j.tplants.2012.02.010>.
- [119] H. Matsui, H. Iwakawa, G.-S. Hyon, I. Yotsui, S. Katou, I. Monte, R. Nishihama, R. Franzen, R. Solano, H. Nakagami, Isolation of natural fungal pathogens from *Marchantia polymorpha* reveals antagonism between salicylic acid and jasmonate during liverwort–fungus interactions, *Plant Cell Physiol.* 61 (2020) 265–275, <https://doi.org/10.1093/pcp/pcz187>.
- [120] R.A. Andersson, M. Akita, M. Pirhonen, E. Gammelgård, J.P.T. Valkonen, Moss-*Erwinia* pathosystem reveals possible similarities in pathogenesis and pathogen defense in vascular and nonvascular plants, *J. Gen. Plant Pathol.* 71 (2005) 23–28, <https://doi.org/10.1007/s10327-004-0154-3>.
- [121] J. Arnerup, M. Nemesio-Goriz, K. Lundén, F.O. Asiegbu, J. Stenlid, M. Elfstrand, The primary module in Norway spruce defence signalling against *H. annosum* s.l. seems to be jasmonate-mediated signalling without antagonism of salicylate-mediated signalling, *Planta* 237 (2013) 1037–1045, <https://doi.org/10.1007/s00425-012-1822-8>.
- [122] S.M. Clarke, S.M. Cristescu, O. Miersch, F.J.M. Harren, C. Wasternack, L.A.J. Mur, Jasmonates act with salicylic acid to confer basal thermotolerance in *Arabidopsis thaliana*, *N. Phytol.* 182 (2009) 175–187, <https://doi.org/10.1111/j.1469-8137.2008.02735.x>.
- [123] M.L. Berens, K.W. Wolinska, S. Spaepen, J. Ziegler, T. Nobori, A. Nair, V. Krüler, T.M. Winkelmüller, Y. Wang, A. Mine, D. Becker, R. Garrido-Oter, P. Schulze-Lefert, K. Tsuda, Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk, *Proc. Natl. Acad. Sci. USA* 116 (2019) 2364–2373, <https://doi.org/10.1073/pnas.1817233116>.
- [124] M. Ke, Z. Ma, D. Wang, Y. Sun, C. Wen, D. Huang, Z. Chen, L. Yang, S. Tan, R. Li, J. Friml, Y. Miao, X. Chen, Salicylic acid regulates *PIN2* auxin transporter hyperclustering and root gravitropic growth via *Remorin*-dependent lipid nanodomain organisation in *Arabidopsis thaliana*, *N. Phytol.* 229 (2021) 963–978, <https://doi.org/10.1111/nph.16915>.
- [125] J. Glazebrook, Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens, *Annu. Rev. Phytopathol.* 43 (2005) 205–227, <https://doi.org/10.1146/annurev.phyto.43.040204.135923>.
- [126] C.M.J. Pieterse, D. Van der Does, C. Zamioudis, A. Leon-Reyes, S.C.M. Van Wees, Hormonal modulation of plant immunity, *Annu. Rev. Cell Dev. Biol.* 28 (2012) 489–521, <https://doi.org/10.1146/annurev-cellbio-092910-154055>.
- [127] C. Ju, B. Van de Poel, E.D. Cooper, J.H. Thierer, T.R. Gibbons, C.F. Delwiche, C. Chang, Conservation of ethylene as a plant hormone over 450 million years of evolution, *Nat. Plants* 1 (2015) 14004, <https://doi.org/10.1038/nplants.2014.4>.
- [128] I. Monte, S. Ishida, A.M. Zamarreño, M. Hamburger, J.M. Franco-Zorrilla, G. García-Casado, C. Gouhier-Darimont, P. Reymond, K. Takahashi, J.M. García-Mina, R. Nishihama, T. Kohchi, R. Solano, Ligand-receptor co-evolution shaped the jasmonate pathway in land plants, *Nat. Chem. Biol.* 14 (2018) 480–488, <https://doi.org/10.1038/s41589-018-0033-4>.
- [129] I. Monte, J.M. Franco-Zorrilla, G. García-Casado, A.M. Zamarreño, J.M. García-Mina, R. Nishihama, T. Kohchi, R. Solano, A single JAZ repressor controls the jasmonate pathway in *Marchantia polymorpha*, *Mol. Plant* 12 (2019) 185–198, <https://doi.org/10.1016/j.molp.2018.12.017>.
- [130] S. Wang, L. Li, H. Li, S.K. Sahu, H. Wang, Y. Xu, W. Xian, B. Song, H. Liang, S. Cheng, Y. Chang, Y. Song, Z. Čebić, S. Wittek, T. Reeder, M. Peterson, H. Yang, J. Wang, B. Melkonian, Y. Van de Peer, X. Xu, G.K.-S. Wong, M. Melkonian, H. Liu, X. Liu, Genomes of early-diverging streptophyte algae shed light on plant terrestrialization, *Nat. Plants* 6 (2020) 95–106, <https://doi.org/10.1038/s41477-019-0560-3>.
- [131] J. Scholz, F. Brodhun, E. Hornung, C. Herrfurth, M. Stumpe, A.K. Beike, B. Faltin, W. Frank, R. Reski, I. Feussner, Biosynthesis of allene oxides in *Physcomitrella patens*, *BMC Plant Biol.* 12 (2012) 228, <https://doi.org/10.1186/1471-2229-12-228>.
- [132] T. Koeduka, K. Ishizaki, C.M. Mwenda, K. Hori, Y. Sasaki-Sekimoto, H. Ohta, T. Kohchi, K. Matsui, Biochemical characterization of allene oxide synthases from the liverwort *Marchantia polymorpha* and green microalgae *Klebsormidium flaccidum* provides insight into the evolutionary divergence of the plant CYP74 family, *Planta* 242 (2015) 1175–1186, <https://doi.org/10.1007/s00425-015-2355-8>.
- [133] M. Stumpe, C. Göbel, B. Faltin, A.K. Beike, B. Hause, K. Himmelsbach, J. Bode, R. Kramell, C. Wasternack, W. Frank, R. Reski, I. Feussner, The moss *Physcomitrella patens* contains cyclopentenones but no jasmonates: mutations in allene oxide cyclase lead to reduced fertility and altered sporophyte morphology, *N. Phytol.* 188 (2010) 740–749, <https://doi.org/10.1111/j.1469-8137.2010.03406.x>.
- [134] P. Pratiwi, G. Tanaka, T. Takahashi, X. Xie, K. Yoneyama, H. Matsuura, K. Takahashi, Identification of jasmonic acid and jasmonoyl-isoleucine, and characterization of AOS, AOC, OPR and JAR1 in the model lycophyte *Selaginella moellendorffii*, *Plant Cell Physiol.* 58 (2017) 789–801, <https://doi.org/10.1093/pcp/pcx031>.
- [135] P.E. Staswick, I. Tiriyaki, The oxylipin signal jasmonic acid is activated by an enzyme that conjugates it to isoleucine in *Arabidopsis*, *Plant Cell* 16 (2004) 2117–2127, <https://doi.org/10.1105/tpc.104.023549>.
- [136] M.J. Beilby, C.E. Turi, T.C. Baker, F.J. Tymms, S.J. Murch, Circadian changes in endogenous concentrations of indole-3-acetic acid, melatonin, serotonin, abscisic acid and jasmonic acid in Characeae (*Chara australis* Brown), *Plant Signal. Behav.* 10 (2015), e1082697, <https://doi.org/10.1080/15592324.2015.1082697>.
- [137] M.S. Gachet, A. Schubert, S. Calarco, J. Boccard, J. Gertsch, Targeted metabolomics shows plasticity in the evolution of signaling lipids and uncovers old and new endocannabinoids in the plant kingdom, *Sci. Rep.* 7 (2017) 41177, <https://doi.org/10.1038/srep41177>.
- [138] M. Peñuelas, I. Monte, F. Schweizer, A. Vallat, P. Reymond, G. García-Casado, J.M. Franco-Zorrilla, R. Solano, Jasmonate-related MYC transcription factors are functionally conserved in *Marchantia polymorpha*, *Plant Cell* 31 (2019) 2491–2509, <https://doi.org/10.1105/tpc.18.00974>.
- [139] I. Monte, S. Kneeshaw, J.M. Franco-Zorrilla, A. Chini, A.M. Zamarreño, J.M. García-Mina, R. Solano, An Ancient CO11-Independent Function for Reactive Electrophilic Oxylipins in Thermotolerance, *Curr. Biol.* 30 (2020) 962–971.e3, <https://doi.org/10.1016/j.cub.2020.01.023>.
- [140] Z. Liang, Y. Geng, C. Ji, H. Du, C.E. Wong, Q. Zhang, Y. Zhang, P. Zhang, A. Riaz, S. Chachar, Y. Ding, J. Wen, Y. Wu, M. Wang, H. Zheng, Y. Wu, V. Demko, L. Shen, X. Han, P. Zhang, X. Gu, H. Yu, *Mesostigma viride* genome and transcriptome provide insights into the origin and evolution of Streptophyta, *1901850–18*, *Adv. Sci.* 38 (2019), <https://doi.org/10.1002/adv.201901850>.
- [141] X. Pu, X. Dong, Q. Li, Z. Chen, L. Liu, An update on the function and regulation of methylerythritol phosphate and mevalonate pathways and their evolutionary dynamics, *J. Integr. Plant Biol.* 63 (2021) 1211–1226, <https://doi.org/10.1111/jipb.13076>.
- [142] E. Vranová, D. Coman, W. Gruißem, Structure and dynamics of the isoprenoid pathway network, *Mol. Plant* 5 (2012) 318–333, <https://doi.org/10.1093/mp/ssp015>.
- [143] J. Lombard, D. Moreira, Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life, *Mol. Biol. Evol.* 28 (2011) 87–99, <https://doi.org/10.1093/molbev/msq177>.
- [144] D.M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics, *Genome Biol.* 20 (2019) 238, <https://doi.org/10.1186/s13059-019-1832-y>.
- [145] N. Nisar, L. Li, S. Lu, N.C. Khin, B.J. Pogson, Carotenoid metabolism in plants, *Mol. Plant* 8 (2015) 68–82, <https://doi.org/10.1016/j.molp.2014.12.007>.
- [146] M. Rippin, B. Becker, A. Holzinger, Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics, *Plant Cell Physiol.* 58 (2017) 2067–2084, <https://doi.org/10.1093/pcp/pcx136>.
- [147] M. Rippin, M. Pichrtová, E. Arc, I. Kranner, B. Becker, A. Holzinger, Metatranscriptomic and metabolite profiling reveals vertical heterogeneity within a *Zygnema* green algal mat from Svalbard (High Arctic), *Environ. Microbiol.* 21 (2019) 4283–4299, <https://doi.org/10.1111/1462-2920.14788>.
- [148] J.M.R. Fürst-Jansen, S. von Vries, M. Lorenz, K. von Schwartzenberg, J. M. Archibald, J. de Vries, Submergence of the filamentous Zygnematophyceae *Mougeotia* induces differential gene expression patterns associated with core metabolism and photosynthesis, *Protoplasma* (2021), <https://doi.org/10.1007/s00709-021-01730-1>.
- [149] J. de Vries, B.A. Curtis, S.B. Gould, J.M. Archibald, Embryophyte stress signaling evolved in the algal progenitors of land plants, *Proc. Natl. Acad. Sci. USA*. 115 (2018) E3471–E3480, <https://doi.org/10.1073/pnas.1719230115>.
- [150] E.-M. Aro, I. Virgin, B. Andersson, Photoinhibition of photosystem II. Inactivation, protein damage and turnover, *Biochim. Biophys. Acta - Bioenerg.* 1143 (1993) 113–134, [https://doi.org/10.1016/0005-2728\(93\)90134-2](https://doi.org/10.1016/0005-2728(93)90134-2).
- [151] P. Müller, X.P. Li, K.K. Niyogi, Non-photochemical quenching. A response to excess light energy, *Plant Physiol.* 125 (2001) 1558–1566, <https://doi.org/10.1104/pp.125.4.1558>.
- [152] R. Goss, B. Lepetit, Biodiversity of NPQ, *J. Plant Physiol.* 172 (2015) 13–32, <https://doi.org/10.1016/j.jplph.2014.03.004>.
- [153] G. Christa, S. Cruz, P. Jahns, J. de Vries, P. Cartaxana, A.C. Esteves, J. Seródio, S. B. Gould, Photoprotection in a monophyletic branch of chlorophyte algae is independent of energy-dependent quenching (qE), *N. Phytol.* 214 (2017) 1132–1144, <https://doi.org/10.1111/nph.14435>.
- [154] V. Giovagnetti, G. Han, M.A. Ware, P. Ungerer, X. Qin, W.-D. Wang, T. Kuang, J.-R. Shen, A.V. Ruban, A siphonous morphology affects light-harvesting modulation in the intertidal green macroalgae *Bryopsis corticulans* (Ulvothyceae), *Planta* 247 (2018) 1293–1306, <https://doi.org/10.1007/s00425-018-2854-5>.
- [155] A. Alboresi, C. Gerotto, G.M. Giacometti, R. Bassi, T. Morosinotto, *Physcomitrella patens* mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization, *Proc. Natl. Acad. Sci. USA* 107 (2010) 11128–11133, <https://doi.org/10.1073/pnas.1002873107>.
- [156] C. Gerotto, A. Alboresi, G.M. Giacometti, R. Bassi, T. Morosinotto, Coexistence of plant and algal energy dissipation mechanisms in the moss *Physcomitrella patens*,

- N. Phytol. 196 (2012) 763–773, <https://doi.org/10.1111/j.1469-8137.2012.04345.x>.
- [157] C. Gerotto, T. Morosinotto, Evolution of photoprotection mechanisms upon land colonization: evidence of PSBS-dependent NPQ in late Streptophyte algae, *Physiol. Plant.* 149 (2013) 583–598, <https://doi.org/10.1111/ppi.12070>.
- [158] V. Correa-Galvis, P. Redekop, K. Guan, A. Griess, T.B. Truong, S. Wakao, K. K. Niyogi, P. Jahns, Photosystem II subunit PsbS is involved in the induction of LHCSR protein-dependent energy dissipation in *Chlamydomonas reinhardtii*, *J. Biol. Chem.* 291 (2016) 17478–17487, <https://doi.org/10.1074/jbc.M116.737312>.
- [159] V. Correa-Galvis, G. Poschmann, M. Melzer, K. Stühler, P. Jahns, PsbS interactions involved in the activation of energy dissipation in *Arabidopsis*, *Nat. Plants* 2 (2016) 1–8, <https://doi.org/10.1038/nplants.2015.225>.
- [160] P. Jahns, D. Latowski, K. Strzalka, Mechanism and regulation of the violaxanthin cycle: the role of antenna proteins and membrane lipids, *Biochim. Biophys. Acta - Bioenerg.* 1787 (2009) 3–14, <https://doi.org/10.1016/j.bbabi.2008.09.013>.
- [161] J. Sacharz, V. Giovagnetti, P. Ungerer, G. Mastroianni, A.V. Ruban, The xanthophyll cycle affects reversible interactions between PsbS and light-harvesting complex II to control non-photochemical quenching, *Nat. Plants* 3 (2017) 16225, <https://doi.org/10.1038/nplants.2016.225>.
- [162] N.E. Holt, D. Zigmantas, L. Valkunas, X.-P. Li, K.K. Niyogi, G.R. Fleming, Carotenoid cation formation and the regulation of photosynthetic light harvesting, *Science* 307 (2005) 433–436, <https://doi.org/10.1126/science.1105833>.
- [163] A.V. Ruban, R. Berera, C. Illoia, I.H.M. van Stokkum, J.T.M. Kennis, A.A. Pascal, H. van Amerongen, B. Robert, P. Horton, R. van Grondelle, Identification of a mechanism of photoprotective energy dissipation in higher plants, *Nature* 450 (2007) 575–578, <https://doi.org/10.1038/nature06262>.
- [164] M. Stamenković, K. Bischof, D. Hanelt, Xanthophyll cycle pool size and composition in several *Cosmarium* strains (Zygnematophyceae, Streptophyta) are related to their geographic distribution patterns, *Protist* 165 (2014) 14–30, <https://doi.org/10.1016/j.protis.2013.10.002>.
- [165] F. Ramel, S. Birtic, S. Cuiñé, C. Triantaphylidés, J.-L. Ravanat, M. Havaux, Chemical quenching of singlet oxygen by carotenoids in plants, *Plant Physiol.* 158 (2012) 1267–1278, <https://doi.org/10.1104/pp.111.182394>.
- [166] X. Hou, J. Rivers, P. León, R.P. McQuinn, B.J. Pogson, Synthesis and function of apocarotenoid signals in plants, *Trends Plant Sci.* 21 (2016) 792–803, <https://doi.org/10.1016/j.tplants.2016.06.001>.
- [167] F. Ramel, S. Birtic, C. Ginies, L. Soubigou-Taconnat, C. Triantaphylidés, M. Havaux, Carotenoid oxidation products are stress signals that mediate gene responses to singlet oxygen in plants, *Proc. Natl. Acad. Sci. USA* 109 (2012) 5535–5540, <https://doi.org/10.1073/pnas.1115982109>.
- [168] J.C. Moreno, J. Mi, Y. Alagoz, S. Al-Babili, Plant apocarotenoids: from retrograde signaling to interspecific communication, *Plant J.* 105 (2021) 351–375, <https://doi.org/10.1111/tpj.15102>.
- [169] K.X. Chan, S.Y. Phua, P. Crisp, R. McQuinn, B.J. Pogson, Learning the languages of the chloroplast: retrograde signaling and beyond, *Annu. Rev. Plant Biol.* 67 (2016) 25–53, <https://doi.org/10.1146/annurev-arplant-043015-111854>.
- [170] K.-P. Jia, J. Mi, S. Ali, H. Ohyanagi, J.C. Moreno, A. Ablazov, A. Balakrishna, L. Berqdar, A. Fiore, G. Diretto, C. Martínez, A.R. de Lera, T. Gojobori, S. Al-Babili, An alternative, zeaxanthin epoxidase-independent abscisic acid biosynthetic pathway in plants, *Mol. Plant* 15 (2022) 151–166, <https://doi.org/10.1016/j.molp.2021.09.008>.
- [171] Y. Sun, Ben Harpazi, A. Wijerathna-Yapa, E. Merilo, J. de Vries, D. Michaeli, M. Gal, A.C. Cuming, H. Kollist, A. Mosquna, A ligand-independent origin of abscisic acid perception, *PNAS* 116 (2019) 24892–24899, <https://doi.org/10.1073/pnas.1914480116>.
- [172] Y. Sun, O. Pri-Tal, D. Michaeli, A. Mosquna, Evolution of abscisic acid signaling module and its perception, *Front. Plant Sci.* 11 (2020) 934, <https://doi.org/10.3389/fpls.2020.00934>.
- [173] K. Mashiguchi, Y. Seto, S. Yamaguchi, Strigolactone biosynthesis, transport and perception, *Plant J.* 105 (2021) 335–350, <https://doi.org/10.1111/tpj.15059>.
- [174] H. Sun, J. Tao, M. Hou, S. Huang, S. Chen, Z. Liang, T. Xie, Y. Wei, X. Xie, K. Yoneyama, G. Xu, Y. Zhang, A strigolactone signal is required for adventitious root formation in rice, *Ann. Bot.* 115 (2015) 1155–1162, <https://doi.org/10.1093/aob/mcv052>.
- [175] C. Ruyter-Spira, W. Kohlen, T. Charnikhova, A. van Zeijl, L. van Bezouwen, N. de Ruijter, C. Cardoso, J.A. Lopez-Raez, R. Matusova, R. Bours, F. Verstappen, H. Bouwmeester, Physiological effects of the synthetic strigolactone analog GR24 on root system architecture in *Arabidopsis*: another belowground role for strigolactones? *Plant Physiol.* 155 (2011) 721–734, <https://doi.org/10.1104/pp.110.166645>.
- [176] V. Gomez-Roldan, S. Fermas, P.B. Brewer, V. Puech-Pagès, E.A. Dun, J.-P. Pillot, F. Letisse, R. Matusova, S. Danoun, J.-C. Portais, H. Bouwmeester, G. Bécard, C. A. Beveridge, C. Rameau, S.F. Rochange, Strigolactone inhibition of shoot branching, *Nature* 455 (2008) 189–194, <https://doi.org/10.1038/nature07271>.
- [177] A. Besserer, V. Puech-Pagès, P. Kiefer, V. Gomez-Roldan, A. Jauneau, S. Roy, J.-C. Portais, C. Roux, G. Bécard, N. Séjalou-Delmas, Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria, *PLoS Biol.* 4 (2006), e226, <https://doi.org/10.1371/journal.pbio.0040226>.
- [178] T. Arite, H. Kameoka, J. Kyoizuka, Strigolactone positively controls crown root elongation in rice, *J. Plant Growth Regul.* 31 (2012) 165–172, <https://doi.org/10.1007/s00344-011-9228-6>.
- [179] K. Akiyama, K. Matsuzaki, H. Hayashi, Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi, *Nature* 435 (2005) 824–827, <https://doi.org/10.1038/nature03608>.
- [180] M.H. Walter, R. Stauder, A. Tissier, Evolution of root-specific carotenoid precursor pathways for apocarotenoid signal biogenesis, *Plant Sci.* 233 (2015) 1–10, <https://doi.org/10.1016/j.plantsci.2014.12.017>.
- [181] M. Wang, M. Schäfer, D. Li, R. Halitschke, C. Dong, E. McGale, C. Paetz, Y. Song, S. Li, J. Dong, S. Heiling, K. Groten, P. Franken, M. Bitterlich, M.J. Harrison, U. Paszkowski, I.T. Baldwin, Blumenols as shoot markers of root symbiosis with arbuscular mycorrhizal fungi, *eLife* 7 (2018), e37093, <https://doi.org/10.7554/eLife.37093>.
- [182] P.-M. Delaux, X. Xie, R.E. Timme, V. Puech-Pagès, C. Dunand, E. Lecompte, C. F. Delwiche, K. Yoneyama, G. Bécard, N. Séjalou-Delmas, Origin of strigolactones in the green lineage, *N. Phytol.* 195 (2012) 857–871, <https://doi.org/10.1111/j.1469-8137.2012.04209.x>.
- [183] M.T. Waters, C. Gutjahr, T. Bennett, D.C. Nelson, Strigolactone signaling and evolution, *Annu. Rev. Plant Biol.* 68 (2017) 291–322, <https://doi.org/10.1146/annurev-arplant-042916-040925>.
- [184] M. Umehara, A. Hanada, S. Yoshida, K. Akiyama, T. Arite, N. Takeda-Kamiya, H. Magome, Y. Kamiya, K. Shirasu, K. Yoneyama, J. Kyoizuka, S. Yamaguchi, Inhibition of shoot branching by new terpenoid plant hormones, *Nature* 455 (2008) 195–200, <https://doi.org/10.1038/nature07272>.
- [185] H. Proust, B. Hoffmann, X. Xie, K. Yoneyama, D.G. Schaefer, K. Yoneyama, F. Nogué, C. Rameau, Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*, *Development* 138 (2011) 1531–1539, <https://doi.org/10.1242/dev.058495>.
- [186] E.L. Decker, A. Alder, S. Hunn, J. Ferguson, M.T. Lehtonen, B. Scheler, K. L. Kerres, G. Wiedemann, V. Safavi-Rizi, S. Nordzieke, A. Balakrishna, L. Baz, J. Avalos, J.P.T. Valkonen, R. Reski, S. Al-Babili, Strigolactone biosynthesis is evolutionarily conserved, regulated by phosphate starvation and contributes to resistance against phytopathogenic fungi in a moss, *Physcomitrella patens*, *N. Phytol.* 216 (2017) 455–468, <https://doi.org/10.1111/nph.14506>.
- [187] K. Kodama, M.K. Rich, A. Yoda, S. Shimazaki, X. Xie, K. Akiyama, Y. Mizuno, A. Komatsu, Y. Luo, H. Suzuki, H. Kameoka, C. Libourel, J. Keller, K. Sakakibara, T. Nishiyama, T. Nakagawa, K. Mashiguchi, K. Uchida, K. Yoneyama, Y. Tanaka, S. Yamaguchi, M. Shimamura, P.-M. Delaux, T. Nomura, J. Kyoizuka, An ancestral function of strigolactones as symbiotic rhizosphere signals, *bioRxiv* (2021), <https://doi.org/10.1101/2021.08.20.457034>.
- [188] P.-M. Delaux, N. Séjalou-Delmas, G. Bécard, J.-M. Ané, Evolution of the plant-microbe symbiotic 'toolkit', *Trends Plant Sci.* 18 (2013) 298–304, <https://doi.org/10.1016/j.tplants.2013.01.008>.
- [189] S. Abe, A. Sado, K. Tanaka, T. Kisugi, K. Asami, S. Ota, H.I. Kim, K. Yoneyama, X. Xie, T. Ohnishi, Y. Seto, S. Yamaguchi, K. Akiyama, K. Yoneyama, T. Nomura, Carlactone is converted to carlactonic acid by MAX1 in *Arabidopsis* and its methyl ester can directly interact with AtD14 in vitro, *Proc. Natl. Acad. Sci. USA* 111 (2014) 18084–18089, <https://doi.org/10.1073/pnas.1410801111>.
- [190] C.H. Walker, K. Siu-Ting, A. Taylor, M.J. O'Connell, T. Bennett, Strigolactone synthesis is ancestral in land plants, but canonical strigolactone signalling is a flowering plant innovation, *BMC Biol.* 17 (2019) 70, <https://doi.org/10.1186/s12915-019-0689-6>.
- [191] Q. Wang, S.M. Smith, J. Huang, Origins of strigolactone and karrikin signaling in plants, *S1360138521003150*, *Trends Plant Sci.* (2021), <https://doi.org/10.1016/j.tplants.2021.11.009>.
- [192] A.J. Simkin, S.H. Schwartz, M. Auldridge, M.G. Taylor, H.J. Klee, The tomato carotenoid cleavage dioxygenase 1 genes contribute to the formation of the flavor volatiles β -ionone, pseudionone, and geranylacetone: carotenoid cleavage dioxygenases of *Lycopersicon esculentum*, *Plant J.* 40 (2004) 882–892, <https://doi.org/10.1111/j.1365-3113.2004.02263.x>.
- [193] L. Shumbe, R. Bott, M. Havaux, Dihydroactinidinolide, a high light-induced β -carotene derivative that can regulate gene expression and photoacclimation in *Arabidopsis*, *Mol. Plant* 7 (2014) 1248–1251, <https://doi.org/10.1093/mp/ssu028>.
- [194] Y. Kotsieridis, R.L. Baumes, A. Bertrand, G.K. Skouroumounis, Quantitative determination of β -ionone in red wines and grapes of Bordeaux using a stable isotope dilution assay, *J. Chromatogr. A* 848 (1999) 317–325, [https://doi.org/10.1016/S0021-9673\(99\)00422-7](https://doi.org/10.1016/S0021-9673(99)00422-7).
- [195] M.M. Mendes-Pinto, Carotenoid breakdown products the-norisoprenoids-in wine aroma, *Arch. Biochem. Biophys.* 483 (2009) 236–245, <https://doi.org/10.1016/j.abb.2009.01.008>.
- [196] J. Langen, P. Wegmann-Herr, H.-G. Schmarr, Quantitative determination of α -ionone, β -ionone, and β -damascenone and enantiomer differentiation of α -ionone in wine for authenticity control using multidimensional gas chromatography with tandem mass spectrometric detection, *Anal. Bioanal. Chem.* 408 (2016) 6483–6496, <https://doi.org/10.1007/s00216-016-9767-6>.
- [197] J.J.B. Timmins, H. Kroukamp, I.T. Paulsen, I.S. Pretorius, The sensory significance of apocarotenoids in wine: importance of Carotenoid Cleavage Dioxygenase 1 (CCD1) in the production of β -ionone, *Molecules* 25 (2020) 2779, <https://doi.org/10.3390/molecules25122779>.
- [198] E. Tomasino, S. Bolman, The potential effect of β -ionone and β -damascenone on sensory perception of Pinot Noir Wine aroma, *Molecules* 26 (2021) 1288, <https://doi.org/10.3390/molecules26051288>.
- [199] L.A. Cáceres, S. Lakshminarayan, K.K.-C. Yeung, B.D. McGarvey, A. Hannoufa, M. W. Sumarah, X. Benitez, I.M. Scott, Repellent and attractive effects of α -, β -, and dihydro- β -ionone to generalist and specialist herbivores, *J. Chem. Ecol.* 42 (2016) 107–117, <https://doi.org/10.1007/s10886-016-0669-z>.
- [200] M. Murata, Y. Nakai, K. Kawazu, M. Ishizaka, H. Kajiwara, H. Abe, K. Takeuchi, Y. Ichinose, I. Mitsuhashi, A. Mochizuki, S. Seo, Loliolide, a carotenoid metabolite, is a potential endogenous inducer of herbivore resistance, *Plant Physiol.* 179 (2019) 1822–1833, <https://doi.org/10.1104/pp.18.00837>.

- [201] K.-I. Harada, K. Ozaki, S. Tsuzuki, H. Kato, M. Hasegawa, E.K. Kuroda, S. Arii, K. Tsuji, Blue color formation of cyanobacteria with β -cyclocitral, *J. Chem. Ecol.* 35 (2009) 1295–1301, <https://doi.org/10.1007/s10886-009-9706-5>.
- [202] X. Liu, C. Shi, X. Xu, X. Li, Y. Xu, H. Huang, Y. Zhao, Y. Zhou, H. Shen, C. Chen, G. Wang, Spatial distributions of β -cyclocitral and β -ionone in the sediment and overlying water of the west shore of Taihu Lake, *Sci. Total Environ.* 579 (2017) 430–438, <https://doi.org/10.1016/j.scitotenv.2016.11.079>.
- [203] S. D'Alessandro, Y. Mizokami, B. Légeret, M. Havaux, The apocarotenoid β -cyclocitric acid elicits drought tolerance in plants, *iScience* 19 (2019) 461–473, <https://doi.org/10.1016/j.isci.2019.08.003>.
- [204] S. Mitra, R. Estrada-Tejedor, D.C. Volke, M.A. Phillips, J. Gershenzon, L.P. Wright, Negative regulation of plastidial isoprenoid pathway by herbivore-induced β -cyclocitral in *Arabidopsis thaliana*, e2008747118, *Proc. Natl. Acad. Sci. USA* 118 (2021), <https://doi.org/10.1073/pnas.2008747118>.
- [205] A.J. Simkin, B.A. Underwood, M. Auldridge, H.M. Louca, K. Shibuya, E. Schmelz, D.G. Clark, H.J. Klee, Circadian regulation of the PhCCD1 carotenoid cleavage dioxygenase controls emission of β -ionone, a fragrance volatile of petunia flowers, *Plant Physiol.* 136 (2004) 3504–3514, <https://doi.org/10.1104/pp.104.049718>.
- [206] J. Mi, K.-P. Jia, A. Balakrishna, Q. Feng, S. Al-Babili, A highly sensitive SPE derivatization-UHPLC-MS approach for quantitative profiling of carotenoid-derived dialdehydes from vegetables, *J. Agric. Food Chem.* 67 (2019) 5899–5907, <https://doi.org/10.1021/acs.jafc.9b01749>.
- [207] K.-P. Jia, A.J. Dickinson, J. Mi, G. Cui, T.T. Xiao, N.M. Kharbatia, X. Guo, E. Sugiono, M. Aranda, I. Bliou, M. Rueping, P.N. Benfey, S. Al-Babili, Anchorene is a carotenoid-derived regulatory metabolite required for anchor root formation in *Arabidopsis*, *Sci. Adv.* 5 (2019) eaaw6787, <https://doi.org/10.1126/sciadv.aaw6787>.
- [208] A.N. Stepanova, J. Robertson-Hoyt, J. Yun, L.M. Benavente, D.-Y. Xie, K. Doležal, A. Schlereth, G. Jürgens, J.M. Alonso, TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development, *Cell* 133 (2008) 177–191, <https://doi.org/10.1016/j.cell.2008.01.047>.
- [209] Y. Zhao, S.K. Christensen, C. Fankhauser, J.R. Cashman, J.D. Cohen, D. Weigel, J. Chory, A role for flavin-monooxygenase-like enzymes in auxin biosynthesis, *Science* 291 (2001) 306–309, <https://doi.org/10.1126/science.291.5502.306>.
- [210] Y. Cheng, X. Dai, Y. Zhao, Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*, *Genes Dev.* 20 (2006) 1790–1799, <https://doi.org/10.1101/gad.1415106>.
- [211] D.M. Eklund, K. Ishizaki, E. Flores-Sandoval, S. Kikuchi, Y. Takebayashi, S. Tsukamoto, Y. Hirakawa, M. Nonomura, H. Kato, M. Kouno, R.P. Bhalarao, U. Lagercrantz, H. Kasahara, T. Kohchi, J.L. Bowman, Auxin produced by the indole-3-pyruvic acid pathway regulates development and gemmae dormancy in the liverwort *Marchantia polymorpha*, *Plant Cell* 27 (2015) 1650–1669, <https://doi.org/10.1105/tpc.15.00065>.
- [212] K. Landberg, J. Šimura, K. Ljung, E. Sundberg, M. Thelander, Studies of moss reproductive development indicate that auxin biosynthesis in apical stem cells may constitute an ancestral function for focal growth control, *N. Phytol.* 229 (2021) 845–860, <https://doi.org/10.1111/nph.16914>.
- [213] S. Kaneko, S.D. Cook, Y. Aoi, A. Watanabe, K.-I. Hayashi, H. Kasahara, An evolutionarily primitive and distinct auxin metabolism in the lycophyte *Selaginella moellendorffii*, *Plant Cell Physiol.* 61 (2020) 1724–1732, <https://doi.org/10.1093/pcp/pcaa098>.
- [214] C. Wang, Y. Liu, S.-S. Li, G.-Z. Han, Origin of plant auxin biosynthesis in charophyte algae, *Trends Plant Sci.* 19 (2014) 741–743, <https://doi.org/10.1016/j.tplants.2014.10.004>.
- [215] F. Romani, Origin of TAA Genes in Charophytes: New Insights into the Controversy over the Origin of Auxin Biosynthesis, *Front. Plant Sci.* 8 (2017) 1616, <https://doi.org/10.3389/fpls.2017.01616>.
- [216] N. Morffy, L.C. Strader, Old Town Roads: routes of auxin biosynthesis across kingdoms, *Curr. Opin. Plant Biol.* 55 (2020) 21–27, <https://doi.org/10.1016/j.pbi.2020.02.002>.
- [217] R. Prusty, P. Grisafi, G.R. Fink, The plant hormone indoleacetic acid induces invasive growth in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. USA* 101 (2004) 4153–4157, <https://doi.org/10.1073/pnas.0400659101>.
- [218] E. Žizková, M. Kubeš, P.I. Dobrev, P. Přibyl, J. Šimura, L. Zahajská, L.Z. Drábková, O. Novák, V. Motyka, Control of cytokinin and auxin homeostasis in cyanobacteria and algae, *Ann. Bot.* 119 (2017) 151–166, <https://doi.org/10.1093/aob/mcw194>.
- [219] K.A. Bogaert, J. Blomme, T. Beeckman, O. De Clerck, Auxin's origin: do PILS hold the key?, S1360138521002570, *Trends Plant Sci.* (2021), <https://doi.org/10.1016/j.tplants.2021.09.008>.
- [220] J. Friml, A. Vieten, M. Sauer, D. Weijers, H. Schwarz, T. Hamann, R. Offringa, G. Jürgens, Efflux-dependent auxin gradients establish the apical-basal axis of *Arabidopsis*, *Nature* 426 (2003) 147–153, <https://doi.org/10.1038/nature02085>.
- [221] I. Bliou, J. Xu, M. Wildwater, V. Willemsen, I. Paponov, J. Friml, R. Heidstra, M. Aida, K. Palme, B. Scheres, The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots, *Nature* 433 (2005) 39–44, <https://doi.org/10.1038/nature03184>.
- [222] M. Adamowski, J. Friml, PIN-dependent auxin transport: action, regulation, and evolution, *Plant Cell* 27 (2015) 20–32, <https://doi.org/10.1105/tpc.114.134874>.
- [223] K.J.M. Boot, K.R. Libbenga, S.C. Hill, R. Offringa, B. Van Duijn, Polar auxin transport: an early invention, *J. Exp. Bot.* 63 (2012) 4213–4218, <https://doi.org/10.1093/jxb/ers106>.
- [224] R. Skokan, E. Medvečková, T. Viaeve, S. Vosolsobě, M. Zwiewka, K. Müller, P. Skůpa, M. Karady, Y. Zhang, D.P. Janacek, U.Z. Hammes, K. Ljung, T. Nodzyński, J. Petrásek, J. Friml, PIN-driven auxin transport emerged early in streptophyte evolution, *Nat. Plants* 5 (2019) 1114–1119, <https://doi.org/10.1038/s41477-019-0542-5>.
- [225] Y. Zhang, L. Rodriguez, L. Li, X. Zhang, J. Friml, Functional innovations of PIN auxin transporters mark crucial evolutionary transitions during rise of flowering plants, *Sci. Adv.* 6 (2020) eabc8895, <https://doi.org/10.1126/sciadv.abc8895>.
- [226] A. Žabka, J.T. Polit, K. Winnicki, P. Paciorek, P. Juszczak, J. Nowak, M. Maszewski, J. PIN2-like proteins may contribute to the regulation of morphogenetic processes during spermatogenesis in *Chara vulgaris*, *Plant Cell Rep.* 35 (2016) 1655–1669, <https://doi.org/10.1007/s00299-016-1979-x>.
- [227] T. Ulmasov, G. Hagen, T.J. Guilfoyle, ARF1, a transcription factor that binds to auxin response elements, *Science* 276 (1997) 1865–1868, <https://doi.org/10.1126/science.276.5320.1865>.
- [228] T. Ulmasov, J. Murfett, G. Hagen, T.J. Guilfoyle, Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements, *Plant Cell* 9 (1997) 1963–1971, <https://doi.org/10.1105/tpc.9.11.1963>.
- [229] W.M. Gray, S. Kepinski, D. Rouse, O. Leyser, M. Estelle, Auxin regulates SCFTIR1-dependent degradation of AUX/IAA proteins 414 (2001) 271–276, <https://doi.org/10.1038/35104500>.
- [230] E. Flores-Sandoval, D.M. Eklund, S.-F. Hong, J.P. Alvarez, T.J. Fisher, E. R. Lampugnani, J.F. Golz, A. Vázquez-Lobo, T. Dierschke, S.-S. Lin, J.L. Bowman, Class C ARFs evolved before the origin of land plants and antagonize differentiation and developmental transitions in *Marchantia polymorpha*, *N. Phytol.* 218 (2018) 1612–1630, <https://doi.org/10.1111/nph.15090>.
- [231] S.K. Mutte, H. Kato, C. Rothfels, M. Melkonian, G.K.-S. Wong, D. Weijers, Origin and evolution of the nuclear auxin response system, *Elife* 7 (2018), <https://doi.org/10.7554/eLife.33399>.
- [232] R. Martin-Arevalillo, E. Thévenon, F. Jégu, T. Vinos-Poyo, T. Vernoux, F. Parcy, R. Dumas, Evolution of the Auxin Response Factors from charophyte ancestors, e1008400-16, *PLoS Genet* 15 (2019), e1008400, <https://doi.org/10.1371/journal.pgen.1008400>.
- [233] B. Van de Poel, E.D. Cooper, D. Van Der Straeten, C. Chang, C.F. Delwiche, Transcriptome Profiling of the Green Alga *Spirogyra pratensis* (Charophyta) Suggests an Ancestral Role for Ethylene in Cell Wall Metabolism, Photosynthesis, and Abiotic Stress Responses, *Plant Physiol.* 172 (2016) 533–545, <https://doi.org/10.1104/pp.16.00299>.
- [234] J.L. Bowman, L.N. Briginshaw, T.J. Fisher, E. Flores-Sandoval, Something ancient and something neofunctionalized—evolution of land plant hormone signaling pathways, *Curr. Opin. Plant Biol.* 47 (2019) 64–72, <https://doi.org/10.1016/j.pbi.2018.09.009>.
- [235] S. Ohno, Evolution by Gene Duplication, Springer, Berlin, Heidelberg, 1970.
- [236] M. Lynch, J.S. Conery, The evolutionary fate and consequences of duplicate genes (<https://www.science.org/doi/>), *Science* 290 (2000) 1151–1155, <https://doi.org/10.1126/science.290.5494.1151>.
- [237] L. Carretero-Paulet, Y. Van de Peer, The evolutionary conundrum of whole-genome duplication, *Am. J. Bot.* 107 (2020) 1101–1105, <https://doi.org/10.1002/ajb2.1520>.
- [238] S.A. Rensing, J. Ick, J.A. Fawcett, D. Lang, A. Zimmer, Y. Van de Peer, R. Reski, An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*, *BMC Evol. Biol.* 7 (2007) 130, <https://doi.org/10.1186/1471-2148-7-130>.
- [239] A. Jozwiak, P.D. Sonawane, S. Panda, C. Garagounis, K.K. Papadopoulou, B. Abebie, H. Massalha, E. Almekias-Siegl, T. Scherf, A. Aharoni, Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery, *Nat. Chem. Biol.* 16 (2020) 740–748, <https://doi.org/10.1038/s41589-020-0541-x>.
- [240] J.-Q. Huang, X. Fang, X. Tian, P. Chen, J.-L. Lin, X.-X. Guo, J.-X. Li, Z. Fan, W.-M. Song, F.-Y. Chen, R. Ahati, L.-J. Wang, Q. Zhao, C. Martin, X.-Y. Chen, Aromatization of natural products by a specialized detoxification enzyme, *Nat. Chem. Biol.* 16 (2020) 250–256, <https://doi.org/10.1038/s41589-019-0446-8>.
- [241] R. Huang, A.J. O'Donnell, J.J. Barboline, T.J. Barkman, Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes, *PNAS* 113 (2016) 10613–10618, <https://doi.org/10.1073/pnas.1602575113>.
- [242] M. Kaltenbach, J.R. Burke, M. Dindo, A. Pabis, F.S. Munsberg, A. Rabin, S.C. L. Kamerlin, J.P. Noel, D.S. Tawfik, Evolution of chalcone isomerase from a noncatalytic ancestor, *Nat. Chem. Biol.* 14 (2018) 548–555, <https://doi.org/10.1038/s41589-018-0042-3>.
- [243] D. Wickell, L.-Y. Kuo, H.-P. Yang, A. Dhabalia Ashok, I. Irisarri, A. Dadras, S. de Vries, J. de Vries, Y.-M. Huang, Z. Li, M.S. Barker, N.T. Hartwick, T.P. Michael, F.-W. Li, Underwater CAM photosynthesis elucidated by Isoetes genome, *Nat. Commun.* 12 (2021) 6348, <https://doi.org/10.1038/s41467-021-26644-7>.
- [244] S. Martens, G. Forkmann, U. Matern, R. Lukačič, Cloning of parsley flavone synthase I, *Phytochemistry* 58 (2001) 43–46, [https://doi.org/10.1016/S0031-9422\(01\)00191-1](https://doi.org/10.1016/S0031-9422(01)00191-1).
- [245] C. Zubieta, X.-Z. He, R.A. Dixon, J.P. Noel, Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases, *Nat. Struct. Biol.* 8 (2001) 271–279, <https://doi.org/10.1038/85029>.
- [246] F. Jacob, Evolution and tinkering, *Science* 196 (1977) 1161–1166, <https://doi.org/10.1126/science.860134>.
- [247] J.A. Raven, M. Giordano, J. Beardall, S.C. Maberly, Algal evolution in relation to atmospheric CO₂: carboxylases, carbon-concentrating mechanisms and carbon oxidation cycles, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367 (2012) 493–507, <https://doi.org/10.1098/rstb.2011.0212>.
- [248] C.A. Juan, J.M. Pérez de la Lastra, F.J. Plou, E. Pérez-Lebeña, The Chemistry of reactive oxygen species (ros) revisited: outlining their role in biological

- macromolecules (DNA, Lipids and Proteins) and induced pathologies, *Int. J. Mol. Sci.* 22 (2021) 4642, <https://doi.org/10.3390/ijms22094642>.
- [249] L. Yang, N. Mih, A. Anand, J.H. Park, J. Tan, J.T. Yurkovich, J.M. Monk, C. J. Lloyd, T.E. Sandberg, S.W. Seo, D. Kim, A.V. Sastry, P. Phaneuf, Y. Gao, J. T. Broddrick, K. Chen, D. Heckmann, R. Szubin, Y. Hefner, A.M. Feist, B. O. Palsson, Cellular responses to reactive oxygen species are predicted from molecular mechanisms, *Proc. Natl. Acad. Sci. USA* 116 (2019) 14368–14373, <https://doi.org/10.1073/pnas.1905039116>.
- [250] K. Das, A. Roychoudhury, Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants, *Front. Environ. Sci.* (2014) 2, <https://doi.org/10.3389/fenvs.2014.00053>.
- [251] S. Steiger, T. Schmitt, H.M. Schaefer, The origin and dynamic evolution of chemical information transfer, *Proc. R. Soc. B* 278 (2011) 970–979, <https://doi.org/10.1098/rspb.2010.2285>.
- [252] J. de Vries, S.A. Rensing, Gene gains paved the path to land, *Nat. Plants* 6 (2020) 7–8, <https://doi.org/10.1038/s41477-019-0579-5>.
- [253] K. Vannerum, M.J.J. Huysman, R. De Rycke, M. Vuylsteke, F. Leliaert, J. Pollier, U. Lütz-Meindl, J. Gillard, L. De Veylder, A. Goossens, D. Inzé, W. Vyverman, Transcriptional analysis of cell growth and morphogenesis in the unicellular green alga *Micrasterias* (Streptophyta), with emphasis on the role of expansin. *BMC Plant Biol.* 11 (2011) 128, <https://doi.org/10.1186/1471-2229-11-128>.
- [254] H. Zhou, A. Wilkens, D. Hanelt, K. von Schwartzberg, Expanding the molecular toolbox for Zygnematophyceae – transient genetic transformation of the desmid *Micrasterias radians* var. *evoluta*, *Eur. J. Phycol.* 56 (2021) 51–60, <https://doi.org/10.1080/09670262.2020.1768298>.
- [255] H. Zhou, K. von Schwartzberg, Zygnematophyceae: from living algae collections to the establishment of future models, *J. Exp. Bot.* 71 (2020) 3296–3304, <https://doi.org/10.1093/jxb/eraa091>.
- [256] I. Sørensen, Z. Fei, A. Andreas, W.G.T. Willats, D.S. Domozych, J.K.C. Rose, Stable transformation and reverse genetic analysis of *Penium margaritaceum*: A platform for studies of charophyte green algae, the immediate ancestors of land plants, *Plant J.* 77 (2014) 339–351, <https://doi.org/10.1111/tjp.12375>.
- [257] M. Regensdorff, M. Deckena, M. Stein, A. Borchers, G. Scherer, M. Lammers, R. Hänsch, S. Zachgo, H. Buschmann, Transient genetic transformation of *Mougeotia scalaris* (Zygnematophyceae) mediated by the endogenous α -tubulin1 promoter, *J. Phycol.* 54 (2018) 840–849, <https://doi.org/10.1111/jpy.12781>.
- [258] N. Cannell, D.M. Emms, A.J. Hetherington, J. MacKay, S. Kelly, L. Dolan, L. J. Sweetlove, Multiple metabolic innovations and losses are associated with major transitions in land plant evolution, *Curr. Biol.* 30 (2020) 1783–1800.e11, <https://doi.org/10.1016/j.cub.2020.02.086>.
- [259] B.A. Halkier, J. Gershenzon, Biology and biochemistry of glucosinolates, *Annu Rev. Plant Biol.* 57 (2006) 303–333, <https://doi.org/10.1146/annurev.arplant.57.032905.105228>.
- [260] S. Zhang, D.A. Bryant, The tricarboxylic acid cycle in cyanobacteria, *Science* 334 (2011) 1551–1553, <https://doi.org/10.1126/science.1210858>.
- [261] S.Y. Rhee, M. Mutwil, Towards revealing the functions of all genes in plants, *Trends Plant Sci.* 19 (2014) 212–221, <https://doi.org/10.1016/j.tplants.2013.10.006>.
- [262] C. Ruprecht, S. Proost, M. Hernandez-Coronado, C. Ortiz-Ramirez, D. Lang, S. A. Rensing, J.D. Becker, K. Vandepoele, M. Mutwil, Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules, *Plant J.* 90 (2017) 447–465, <https://doi.org/10.1111/tjp.13502>.
- [263] F. Chen, A. Ludwiczuk, G. Wei, X. Chen, B. Crandall-Stotler, J.L. Bowman, Terpenoid secondary metabolites in bryophytes: chemical diversity, biosynthesis and biological functions, *Crit. Rev. Plant Sci.* 37 (2018) 210–231, <https://doi.org/10.1080/07352689.2018.1482397>.
- [264] J. de Vries, T. Ischebeck, Ties between stress and lipid droplets pre-date seeds, *Trends Plant Sci.* 25 (2020) 1203–1214, <https://doi.org/10.1016/j.tplants.2020.07.017>.
- [265] F. Romani, E. Banić, S.N. Florent, T. Kanazawa, J.Q.D. Goodger, R.A. Mentink, T. Dierschke, S. Zachgo, T. Ueda, J.L. Bowman, M. Tsiantis, J.E. Moreno, Oil body formation in *Marchantia polymorpha* is controlled by MpC1HDZ and serves as a defense against arthropod herbivores, *Curr. Biol.* 30 (2020) 2815–2828.e8, <https://doi.org/10.1016/j.cub.2020.05.081>.
- [266] C. Lemieux, C. Otis, M. Turmel, A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atrophycicus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies, *BMC Biol.* 5 (2007) 2, <https://doi.org/10.1186/1741-7007-5-2>.
- [267] J. de Vries, A. Stanton, J.M. Archibald, S.B. Gould, Streptophyte terrestrialization in light of plastid evolution, *Trends Plant Sci.* 21 (2016) 467–476, <https://doi.org/10.1016/j.tplants.2016.01.021>.
- [268] N.J. Wickett, S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, M.S. Barker, J.G. Burleigh, M.A. Gitzendanner, B.R. Ruhfel, E. Wafila, J.P. Der, S.W. Graham, S. Mathews, M. Melkonian, D.E. Soltis, P. S. Soltis, N.W. Miles, C.J. Rothfels, L. Pokorny, A.J. Shaw, L. De Gironimo, D. W. Stevenson, B. Surek, J.C. Villarreal, B. Roure, H. Philippe, C.W. De Pamphilis, T. Chen, M.K. Deyholos, R.S. Baucom, T.M. Kutchan, M.M. Augustin, J. Wang, Y. Zhang, Z. Tian, Z. Yan, X. Wu, X. Sun, G.K.-S. Wong, J. Leebens-Mack, Phylotranscriptomic analysis of the origin and early diversification of land plants, *Proc. Natl. Acad. Sci. USA* 111 (2014) E4859–E4868, <https://doi.org/10.1073/pnas.1323926111>.
- [269] S. Wodniok, H. Brinkmann, G. Glockner, A. Heidel, H. Philippe, M. Melkonian, B. Becker, Origin of land plants: Do conjugating green algae hold the key? *BMC Evol. Biol.* 11 (2011) 104, <https://doi.org/10.1186/1471-2148-11-104>.
- [270] M.N. Puttick, J.L. Morris, T.A. Williams, C.J. Cox, D. Edwards, P. Kenrick, S. Pressel, C.H. Wellman, H. Schneider, D. Pisani, P.C.J. Donoghue, The interrelationships of land plants and the nature of the ancestral embryophyte, *Curr. Biol.* 28 (2018) 733–745, <https://doi.org/10.1016/j.cub.2018.01.063>.
- [271] C.E. Rogers, K.R. Mattox, K.D. Stewart, The zoospore of *Chlorokybus atrophycicus*, a charophyte with sarcinoid growth habit, *Am. J. Bot.* 67 (1980) 774–783, <https://doi.org/10.1002/j.1537-2197.1980.tb07706.x>.
- [272] G.M. Lokhorst, H.J. Sluiman, W. Star, The ultrastructure of mitosis and cytokinesis in the sarcinoid *Chlorokybus atrophycicus* (Chlorophyta, Charophyceae) revealed by rapid freeze fixation and freeze substitution, *J. Phycol.* 24 (1988) 237–248, <https://doi.org/10.1111/j.1529-8817.1988.tb04239.x>.
- [273] R.M. McCourt, C.F. Delwiche, K.G. Karol, Charophyte algae and land plant origins, *Ecol. Evol.* 19 (2004) 661–666, <https://doi.org/10.1016/j.tree.2004.09.013>.
- [274] A.A. Gontcharov, M. Melkonian, Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and rbcL sequence comparisons, *Phycologia* 43 (2004) 105–113, <https://doi.org/10.21216/10031-8884-43-1-105.1>.
- [275] I. Irisarri, T. Darienko, T. Pröschold, J.M.R. Fürst-Jansen, M. Jamy, J. de Vries, Unexpected cryptic species among streptophyte algae most distant to land plants, 20212168, *Proc. R. Soc. B* 288 (2021), 20212168, <https://doi.org/10.1098/rspb.2021.2168>.
- [276] K.G. Karol, R.M. McCourt, M.T. Cimino, C.F. Delwiche, The closest living relatives of land plants, *Science* 294 (2001) 2351–2353, <https://doi.org/10.1126/science.1065156>.
- [277] M. Turmel, M. Ehara, C. Otis, C. Lemieux, Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences, *J. Phycol.* 38 (2002) 364–375, <https://doi.org/10.1046/j.1529-8817.2002.01163.x>.
- [278] E. Cocquyt, H. Verbruggen, F. Leliaert, O. De Clerck, Evolution and cytological diversification of the green seaweeds (Ulvophyceae), *Mol. Biol. Evol.* 27 (2010) 2052–2061, <https://doi.org/10.1093/molbev/msq091>.
- [279] C. Finet, R.E. Timme, C.F. Delwiche, F. Marlétaz, Multigene phylogeny of the green lineage reveals the origin and diversification of land plants, *Curr. Biol.* 20 (2010) 2217–2222, <https://doi.org/10.1016/j.cub.2010.11.035>.
- [280] R.M. McCourt, K.G. Karol, J.D. Hall, M.T. Casanova, M.C. Grant, Charophyceae (Charales), in: L. Margulis, J. Archibald, A. Simpson, C. Slamovits (Eds.), *Handbook of the Protists*, Springer International Publishing, 2017, pp. 165–183.
- [281] F. Leliaert, D.R. Smith, H. Moreau, M. Herron, H. Verbruggen, C.F. Delwiche, O. De Clerck, Phylogeny and molecular evolution of the green algae, *Crit. Rev. Plant Sci.* 31 (2012) 1–46, <https://doi.org/10.1080/07352689.2011.615705>.
- [282] H. Lu, I. Soulié-Marsché, Q.E. Wang, Evolution and classification of Palaeozoic charophytes, *Acta Micro Sin.* 13 (1996) 1–12.
- [283] M. Feist, J. Liu, P. Tafforeau, New insights into Paleozoic charophyte morphology and phylogeny, *Am. J. Bot.* 92 (2005) 1152–1160, <https://doi.org/10.3732/ajb.92.7.1152>.
- [284] V.W. Proctor, Historical biogeography of Chara (Charophyta): An appraisal of the Braun-Wood classification plus a falsifiable alternative for future consideration, *J. Phycol.* 16 (1980) 218–233, <https://doi.org/10.1111/j.1529-8817.1980.tb03023.x>.
- [285] L.E. Graham, *Coleochaete* and the origin of land plants, *Am. J. Bot.* 71 (1984) 603–608, <https://doi.org/10.2307/2443336>.
- [286] H. Printz, Die Chaetophorales der Binnengewässer, Die Chaetophorales der Binnengewässer: Eine systematische übersicht, *Hydrobiologia* 24 (1964) 1–376, <https://doi.org/10.1007/BF00170411>.
- [287] H.J. Marchant, J.D. Pickett-Heaps, Mitosis and cytokinesis in *Coleochaete scutata*, *J. Phycol.* 9 (1973) 461–471, <https://doi.org/10.1111/j.1529-8817.1973.tb04122.x>.
- [288] L.E. Graham, M.E. Cook, J.S. Busse, The origin of plants: body plan changes contributing to a major evolutionary radiation, *Proc. Natl. Acad. Sci. USA* 97 (2000) 4535–4540, <https://doi.org/10.1073/pnas.97.9.4535>.
- [289] M.E. Cook, Cytokinesis in *Coleochaete orbicularis* (Charophyceae): an ancestral mechanism inherited by plants, *Am. J. Bot.* 91 (2004) 313–320, <https://doi.org/10.3732/ajb.91.3.313>.
- [290] C.F. Delwiche, K.G. Karol, M.T. Cimino, K.J. Sytsma, Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene rbcL, *J. Phycol.* 38 (2002) 394–403, <https://doi.org/10.1046/j.1529-8817.2002.01174.x>.
- [291] J.D. Hall, C.F. Delwiche, In the shadow of giants: systematics of the charophyte green algae, in: J. Brodie, J. Lewis (Eds.), *Unravelling the algae: The past, present, and future of algal systematics*, CRC Press, Boca Raton, FL, 2007, pp. 155–169. ISBN 9780367388195.
- [292] R.M. McCourt, K.G. Karol, J. Bell, K.M. Helm-Bychowski, A. Grajewska, M. F. Wojciechowski, R.W. Hoshaw, Phylogeny of the conjugating green algae (Zygnematophyceae) based on rbcL sequences, *J. Phycol.* 36 (2000) 747–758, <https://doi.org/10.1046/j.1529-8817.2000.99106.x>.
- [293] A.A. Gontcharov, B. Marin, M. Melkonian, Molecular phylogeny of conjugating green algae (Zygnematophyceae, Streptophyta) inferred from SSU rDNA sequence comparisons, *J. Mol. Evol.* 56 (2003) 89–104, <https://doi.org/10.1007/s00239-002-2383-4>.
- [294] A.A. Gontcharov, B. Marin, M. Melkonian, Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and rbcL sequence comparisons in the Zygnematophyceae (Streptophyta), *Mol. Biol. Evol.* 21 (2004) 612–624, <https://doi.org/10.1093/molbev/msh052>.
- [295] A.A. Gontcharov, M. Melkonian, In search of monophyletic taxa in the family Desmidiaceae (Zygnematophyceae, Viridiplantae): the genus *Cosmarium*, *Am. J. Bot.* 95 (2008) 1079–1095, <https://doi.org/10.3732/ajb.0800046>.

- [296] J.D. Hall, R.M. McCourt, Zygnematophyta, in: L. Margulis, J. Archibald, A. Simpson, C. Slamovits (Eds.), *Handbook of the Protists*, Springer International Publishing, 2017, pp. 135–163.
- [297] J.Z. Kadlubowska, Conjugatophyceae I – Zygnematales, in: H. Ettl, H. Gerloff, H. Heynig, D. Mollenhauer (Eds.), *Süßwasserflora von Mitteleuropa, Chlorophyta VIII*, Gustav Fischer Verlag, Stuttgart, New York, 1984.
- [298] E. Kol, Kryobiologie. Biologie und Limnologie des Schnees und Eises. I. Kryovegetation, Binnengewässer 24, Schweizerbart Stuttgart (1968) 216 p, ISBN 978-3-510-40033-1.
- [299] B. Büdel, T. Darienko, K. Deuschewitz, S. Dojani, T. Friedl, K.I. Mohr, M. Salisch, W. Reisser, B. Weber, Southern African biological soil crusts are ubiquitous and highly diverse in drylands, being restricted by rainfall frequency, *Micro Ecol.* 57 (2009) 229–247, <https://doi.org/10.1007/s00248-008-9449-9>.
- [300] L.A.A. Zettler, F. Gomez, E. Zettler, B.G. Keenan, R. Amils, M.L. Sogin, Eukaryotic diversity in Spain's River of Fire, *Nature* 417 (2002) 137, <https://doi.org/10.1038/417137a>.
- [301] A. Holzinger, A. Tschalkner, D. Remias, Cytoarchitecture of the desiccation-tolerant green alga *Zygodonium ericetorum*, *Protoplasma* 243 (2010) 15–24, <https://doi.org/10.1007/s00709-009-0048-5>.
- [302] J. Růžicka, Die Desmidiaceen Mitteleuropas. Band 1 (1977) 1. Lieferung. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 291 pp. ISBN 978-3-510-65078-1.
- [303] M. Mix, Die Feinstruktur der Zellwände bei Mesotaeniaceae und Gonatozygaceae mit einer vergleichenden Betrachtung der verschiedenen Wandentypen der Conjugatophyceae und über deren systematischen Wert, *Arch. Mikrobiol.* 81 (1972) 197–220, <https://doi.org/10.1007/BF00412239>.
- [304] J.D. Hall, K.G. Karol, R.M. McCourt, C.F. Delwiche, Phylogeny of the conjugating green algae based on chloroplast and mitochondrial nucleotide sequence data, *J. Phycol.* 44 (2008) 467–477, <https://doi.org/10.1111/j.1529-8817.2008.00485.x>.
- [305] T. Tohge, M. Watanabe, R. Hoefgen, A. Fernie, Shikimate and phenylalanine biosynthesis in the green lineage, *Front. Plant Sci.* 4 (2013) 62, <https://doi.org/10.3389/fpls.2013.00062>.
- [306] T. Tohge, M. Watanabe, R. Hoefgen, A.R. Fernie, The evolution of phenylpropanoid metabolism in the green lineage, *Crit. Rev. Biochem. Mol. Biol.* 48 (2013) 123–152, <https://doi.org/10.3109/10409238.2012.758083>.
- [307] F.-W. Li, P. Brouwer, L. Carretero-Paulet, S. Cheng, J. de Vries, P.-M. Delaux, A. Eily, N. Koppers, L.-Y. Kuo, Z. Li, M. Simenc, I. Small, E. Wafula, S. Angarita, M.S. Barker, A. Brütigam, C. dePamphilis, S. Gould, P.S. Hosmani, Y.-M. Huang, B. Huettel, Y. Kato, X. Liu, S. Maere, R. McDowell, L.A. Mueller, K.G.J. Nierop, S. A. Rensing, T. Robison, C.J. Rothfels, E.M. Sigel, Y. Song, P.R. Timilsena, Y. Van de Peer, H. Wang, P.K.I. Wilhelmsson, P.G. Wolf, X. Xu, J.P. Der, H. Schluepmann, G.K.-S. Wong, K.M. Pryer, Fern genomes elucidate land plant evolution and cyanobacterial symbioses, *Nat. Plants* 4 (2018) 460–472, <https://doi.org/10.1038/s41477-018-0188-8>.
- [308] S. de Vries, J. de Vries, H. Teschke, J.K. von Dahlen, L.E. Rose, S.B. Gould, Jasmonic and salicylic acid response in the fern *Azolla filiculoides* and its cyanobiont, *Plant, Cell Environ.* 41 (2018) 2530–2548, <https://doi.org/10.1111/pce.13131>.
- [309] L. Kriegshauser, S. Knosp, E. Grienenberger, K. Tatsumi, D.D. Gütle, I. Sørensen, L. Herrgott, J. Zumsteg, J.K.C. Rose, R. Reski, D. Werck-Reichhart, H. Renault, Function of the HYDROXYCINNAMOYL-CoA:SHIKIMATE HYDROXYCINNAMOYL TRANSFERASE is evolutionarily conserved in embryophytes, *Plant Cell* 33 (2021) 1472–1491, <https://doi.org/10.1093/plcell/koab044>.
- [310] V.A. Albert, W.B. Barbazuk, C.W. dePamphilis, J.P. Der, J. Leebens-Mack, H. Ma, J.D. Palmer, S. Rounsley, D. Sankoff, S.C. Schuster, D.E. Soltis, P.S. Soltis, S. R. Wessler, R.A. Wing, V.A. Albert, J.S.S. Ammiraju, W.B. Barbazuk, S. Chamala, A.S. Chanderbali, C.W. dePamphilis, J.P. Der, R. Determann, J. Leebens-Mack, H. Ma, P. Ralph, S. Rounsley, S.C. Schuster, D.E. Soltis, P.S. Soltis, J. Talag, L. Tomsho, B. Walts, S. Wanke, R.A. Wing, V.A. Albert, W.B. Barbazuk, S. Chamala, A.S. Chanderbali, T.-H. Chang, R. Determann, T. Lan, D.E. Soltis, P. S. Soltis, S. Arikiti, M.J. Axtell, S. Ayyampalayam, W.B. Barbazuk, I.I.I. James, M. Burnette, S. Chamala, E.D. Paoli, C.W. dePamphilis, J.P. Der, J.C. Estill, N. P. Farrell, A. Harkess, Y. Jiao, J. Leebens-Mack, K. Liu, W. Mei, B.C. Meyers, S. Shahid, E. Wafula, B. Walts, S.R. Wessler, J. Zhai, X. Zhang, V.A. Albert, L. Carretero-Paulet, C.W. dePamphilis, J.P. Der, Y. Jiao, J. Leebens-Mack, E. Lyons, D. Sankoff, H. Tang, E. Wafula, C. Zheng, V.A. Albert, N.S. Altman, W. B. Barbazuk, L. Carretero-Paulet, C.W. dePamphilis, J.P. Der, J.C. Estill, Y. Jiao, J. Leebens-Mack, K. Liu, W. Mei, E. Wafula, N.S. Altman, S. Arikiti, M.J. Axtell, S. Chamala, A.S. Chanderbali, F. Chen, J.-Q. Chen, Y. Chiang, E.D. Paoli, C. W. dePamphilis, J.P. Der, R. Determann, B. Fogliani, C. Guo, J. Harholt, A. Harkess, C. Job, D. Job, S. Kim, H. Kong, J. Leebens-Mack, G. Li, L. Li, J. Liu, H. Ma, B.C. Meyers, J. Park, X. Qi, L. Rajjou, V. Burtet-Sarramegna, R. Sederoff, S. Shahid, D.E. Soltis, P.S. Soltis, Y.-H. Sun, P. Ulvskov, M. Villegente, J.-Y. Xue, T.-F. Yeh, X. Yu, J. Zhai, J.J. Acosta, V.A. Albert, W.B. Barbazuk, R.A. Bruenn, S. Chamala, A. de Kochko, C.W. dePamphilis, J.P. Der, L.R. Herrera-Estrella, E. Ibarra-Laclette, M. Kirst, J. Leebens-Mack, S.P. Pissis, V. Poncet, S.C. Schuster, D.E. Soltis, P.S. Soltis, L. Tomsho, The *Amborella* genome and the evolution of flowering plants, *Science* (2013), <https://doi.org/10.1126/science.1241089>.
- [311] J.A. Banks, T. Nishiyama, M. Hasebe, J.L. Bowman, M. Gribskov, C. dePamphilis, V.A. Albert, N. Aono, T. Aoyama, B.A. Ambrose, N.W. Ashton, M.J. Axtell, E. Barker, M.S. Barker, J.L. Bennetzen, N.D. Bonawitz, C. Chapple, C. Cheng, L.G. G. Correa, M. Dacre, J. DeBarry, I. Dreyer, M. Elias, E.M. Engstrom, M. Estelle, L. Feng, C. Finet, S.K. Floyd, W.B. Frommer, T. Fujita, L. Gramzow, M. Gutensohn, J. Harholt, M. Hattori, A. Heyl, T. Hirai, Y. Hiwatashi, M. Ishikawa, M. Iwata, K. G. Karol, B. Koehler, U. Kolkusaoglu, M. Kubo, T. Kurata, S. Lalonde, K. Li, Y. Li, A. Litt, E. Lyons, G. Manning, T. Maruyama, T.P. Michael, K. Mikami, S. Miyazaki, S. Morinaga, T. Murata, B. Mueller-Roebber, D.R. Nelson, M. Obara, Y. Oguri, R. G. Olinstead, N. Onodera, B.L. Petersen, B. Pils, M. Prigge, S.A. Rensing, D. M. Riaño-Pachón, A.W. Roberts, Y. Sato, H.V. Scheller, B. Schulz, C. Schulz, E. V. Shakhov, N. Shibagaki, N. Shinohara, D.E. Shippen, I. Sørensen, R. Sotooka, N. Sugimoto, M. Sugita, N. Sumikawa, M. Tanurdzic, G. Theißen, P. Ulvskov, S. Wakazuki, J.-K. Weng, W.W.G.T. Willats, D. Wipf, P.G. Wolf, L. Yang, A. D. Zimmer, Q. Zhu, T. Mitros, U. Hellsten, D. Loqué, R. Otitlar, A. Salamov, J. Schmutz, H. Shapiro, E. Lindquist, S. Lucas, D. Rokhsar, I.V. Grigoriev, The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants, *Science* 332 (2011) 960–963, <https://doi.org/10.1126/science.1203810>.
- [312] G. Blanc, I. Agarkova, J. Grimwood, A. Kuo, A. Brueggeman, D.D. Dunigan, J. Gurnon, I. Ladunga, E. Lindquist, S. Lucas, J. Panglinan, T. Pröschold, A. Salamov, J. Schmutz, D. Weeks, T. Yamada, A. Lomsadze, M. Borodovsky, J.-M. Claverie, I.V. Grigoriev, J.L. Van Etten, The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation, *Genome Biol.* 13 (2012) R39, <https://doi.org/10.1186/gb-2012-13-5-r39>.
- [313] T.T. Hu, P. Pattyn, E.G. Bakker, J. Cao, J.-F. Cheng, R.M. Clark, N. Fahlgren, J. A. Fawcett, J. Grimwood, H. Gundlach, G. Haber, J.D. Hollister, S. Ossowski, R. P. Ottillar, A.A. Salamov, K. Schneeberger, M. Spannagl, X. Wang, L. Yang, M. E. Nasrallah, J. Bergelson, J.C. Carrington, B.S. Gaut, J. Schmutz, K.F.X. Mayer, Y. Van de Peer, I.V. Grigoriev, M. Nordborg, D. Weigel, Y.-L. Guo, The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change, *Nat. Genet.* 43 (2011) 476–481, <https://doi.org/10.1038/ng.807>.
- [314] P. Lamesch, T.Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Mueller, K. Dreher, D.L. Alexander, M. Garcia-Hernandez, A.S. Kartikayan, C. H. Lee, W.D. Nelson, L. Ploetz, S. Singh, A. Wensel, E. Huala, The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools, *Nucleic Acids Res.* 40 (2012) D1202–D1210, <https://doi.org/10.1093/nar/gkr1090>.
- [315] D. Lang, K.K. Ullrich, F. Murat, J. Fuchs, J. Jenkins, F.B. Haas, M. Piednoel, H. Gundlach, M. Van Bel, R. Meyberg, C. Vives, J. Morata, A. Symeonidi, M. Hiss, W. Muchero, Y. Kamisugi, O. Saleh, G. Blanc, E.L. Decker, N. van Gessel, J. Grimwood, R.D. Hayes, S.W. Graham, L.E. Gunter, S.F. McDaniel, S.N. W. Hoernstein, A. Larsson, F.-W. Li, P.-F. Perroud, J. Phillips, P. Ranjan, D. S. Rokhsar, C.J. Rothfels, L. Schneider, S. Shu, D.W. Stevenson, F. Thümmel, M. Tillich, J.C. Villarreal Aguilar, T. Widiez, G.K.-S. Wong, A. Wymore, Y. Zhang, A.D. Zimmer, R.S. Quatrano, K.F.X. Mayer, D. Goodstein, J.M. Casacuberta, K. Vandepoel, R. Reski, A.C. Cuming, G.A. Tuskan, F. Maumus, J. Salse, J. Schmutz, S.A. Rensing, The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution, *Plant J.* 93 (2018) 515–533, <https://doi.org/10.1111/tpl.13801>.
- [316] S.S. Merchant, S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, A. Terry, A. Salamov, L.K. Fritz-Laylin, L. Maréchal-Drouard, W.F. Marshall, L.-H. Qu, D.R. Nelson, A.A. Sanderfoot, M.H. Spalding, V.V. Kapitonov, Q. Ren, P. Ferris, E. Lindquist, H. Shapiro, S.M. Lucas, J. Grimwood, J. Schmutz, P. Cardol, H. Cerutti, G. Chanfreau, C.-L. Chen, V. Cognat, M.T. Croft, R. Dent, S. Dutcher, E. Fernández, H. Fukuzawa, D. González-Ballester, D. González-Halphen, A. Hallmann, M. Hanikenne, M. Hippler, W. Inwood, K. Jabbari, M. Kalanon, R. Kuras, P.A. Lefebvre, S.D. Lemaire, A.V. Lobanov, M. Lohr, A. Manuell, I. Meier, L. Mets, M. Mittag, T. Mittelmeier, J.V. Moroney, J. Moseley, C. Napoli, A.M. Nedelcu, K. Niyogi, S.V. Novoselov, I.T. Paulsen, G. Pazour, S. Purton, J.-P. Ral, D.M. Riaño-Pachón, W. Riekhof, L. Rymarquis, M. Schroda, D. Stern, J. Umen, R. Willows, N. Wilson, S.L. Zimmer, J. Allmer, J. Balk, K. Bisova, C.-J. Chen, M. Elias, K. Gendler, C. Hauser, M.R. Lamb, H. Ledford, J. C. Long, J. Minagawa, M.D. Page, J. Pan, W. Pootakham, S. Roje, A. Rose, E. Stahlberg, A.M. Terauchi, P. Yang, S. Ball, C. Bowler, C.L. Dieckmann, V. N. Gladyshev, P. Green, R. Jorgensen, S. Mayfield, B. Mueller-Roebber, S. Rajamani, R.T. Sayre, P. Borkstein, I. Dubchak, D. Goodstein, L. Hornick, Y. W. Huang, J. Jhaveri, Y. Luo, D. Martínez, W.C.A. Ngau, B. Otitlar, A. Poliakov, A. Porter, L. Szajkowski, G. Werner, K. Zhou, I.V. Grigoriev, D.S. Rokhsar, A. R. Grossman, The *Chlamydomonas* genome reveals the evolution of key animal and plant functions, *Science* 318 (2007) 245–250, <https://doi.org/10.1126/science.1143609>.
- [317] B. Nystedt, N.R. Street, A. Wetterbom, A. Zuccolo, Y.-C. Lin, D.G. Scofield, F. Vezzi, N. Delhomme, S. Giacomello, A. Alexeyenko, R. Vicedomini, K. Sahlin, E. Sherwood, M. Elfstrand, L. Gramzow, K. Holmberg, J. Hallman, O. Keech, L. Klason, M. Koriabine, M. Kucukoglu, M. Källér, J. Luthman, F. Lysholm, T. Niitylä, Å. Olson, N. Rilakovic, C. Ritland, J.A. Rosselló, J. Sena, T. Svensson, C. Talavera-López, G. Theißen, H. Tuominen, K. Vanneste, Z.-Q. Wu, B. Zhang, P. Zerbe, L. Arvestad, R. Bhalerao, J. Bohlmann, J. Bousquet, R. Garcia Gil, T. R. Hvidsten, P. de Jong, J. MacKay, M. Morgante, K. Ritland, B. Sundberg, S. Lee Thompson, Y. Van de Peer, B. Andersson, O. Nilsson, P.K. Ingvarsson, J. Lundeberg, S. Jansson, The Norway spruce genome sequence and conifer genome evolution, *Nature* 497 (2013) 579–584, <https://doi.org/10.1038/nature12211>.
- [318] S. Ouyang, W. Zhu, J. Hamilton, H. Lin, M. Campbell, K. Childs, F. Thibaud-Nissen, R.L. Malek, Y. Lee, L. Zheng, J. Orvis, B. Haas, J. Wortman, C.R. Buell, The TIGR Rice Genome Annotation Resource: improvements and new features, *Nucleic Acids Res.* 35 (2007) D883–D887, <https://doi.org/10.1093/nar/gkl976>.
- [319] B. Palenik, J. Grimwood, A. Aerts, P. Rouzé, A. Salamov, N. Putnam, C. Dupont, R. Jorgensen, E. Derelle, S. Rombauts, K. Zhou, R. Otitlar, S.S. Merchant, S. Podell, T. Gaasterland, C. Napoli, K. Gendler, A. Manuell, V. Tai, O. Vallon, G. Piganeau, S. Jancek, M. Heijde, K. Jabbari, C. Bowler, M. Lohr, S. Robbens, G. Werner, I. Dubchak, G.J. Pazour, Q. Ren, I. Paulsen, C. Delwiche, J. Schmutz,

- D. Rokhsar, Y.V. de Peer, H. Moreau, I.V. Grigoriev, The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation, *PNAS* 104 (2007) 7705–7710, <https://doi.org/10.1073/pnas.0611046104>.
- [320] J.P. Vogel, D.F. Garvin, T.C. Mockler, J. Schmutz, D. Rokhsar, M.W. Bevan, K. Barry, S. Lucas, M. Harmon-Smith, K. Lail, H. Tice, J. Schmutz (Leader), J. Grimwood, N. McKenzie, M.W. Bevan, N. Huo, Y.Q. Gu, G.R. Lazo, O.D. Anderson, J.P. Vogel (Leader), F.M. You, M.-C. Luo, J. Dvorak, J. Wright, M. Febrer, M.W. Bevan, D. Idziak, R. Hasterok, D.F. Garvin, E. Lindquist, M. Wang, S. E. Fox, H.D. Priest, S.A. Filichkin, S.A. Givan, D.W. Bryant, J.H. Chang, T.C. Mockler (Leader), H. Wu, W. Wu, A.-P. Hsia, P.S. Schnable, A. Kalyanaraman, B. Barbazuk, T.P. Michael, S.P. Hazen, J.N. Bragg, D. Laudencia-Chingcuanco, J.P. Vogel, D.F. Garvin, Y. Weng, N. McKenzie, M.W. Bevan, G. Haberer, M. Spannagl, K. Mayer (Leader), T. Rattei, T. Mitros, D. Rokhsar, S.-J. Lee, J.K.C. Rose, L.A. Mueller, T.L. York, T. Wicker (Leader), J.P. Buchmann, J. Tanskanen, A.H. Schulman (Leader), H. Gundlach, J. Wright, M. Bevan, A. Costa de Oliveira, L. da C. Maia, Y. Belknap, Y.Q. Gu, N. Jiang, J. Lai, L. Zhu, J. Ma, C. Sun, E. Pritcham, J. Salse (Leader), F. Murat, M. Abrouk, G. Haberer, M. Spannagl, K. Mayer, R. Bruggmann, J. Messing, F.M. You, M.-C. Luo, J. Dvorak, N. Fahlgren, S.E. Fox, C. M. Sullivan, T.C. Mockler, J.C. Carrington, E.J. Chapman, G.D. May, J. Zhai, M. Ganssmann, S. Guna Ranjan Gurazada, M. German, B.C. Meyers, P.J. Green (Leader), J.N. Bragg, L. Tyler, J. Wu, Y.Q. Gu, G.R. Lazo, D. Laudencia-Chingcuanco, J. Thomson, J.P. Vogel (Leader), S.P. Hazen, S. Chen, H.V. Scheller, J. Harholt, P. Ulvskov, S.E. Fox, S.A. Filichkin, N. Fahlgren, J.A. Kimbrel, J.H. Chang, C.M. Sullivan, E.J. Chapman, J.C. Carrington, T.C. Mockler, L.E. Bartley, P. Cao, K.-H. Jung, M.K. Sharma, M. Vega-Sanchez, P. Ronald, C.D. Dardick, S. De Bodt, W. Verelst, D. Inzé, M. Heese, A. Schnittger, X. Yang, U.C. Kalluri, G.A. Tuskan, Z. Hua, R.D. Vierstra, D.F. Garvin, Y. Cui, S. Ouyang, Q. Sun, Z. Liu, A. Yilmaz, E. Grotewold, R. Sibout, K. Hematy, G. Mouille, H. Höfte, T. Michael, J. Pelloux, D. O'Connor, J. Schnable, S. Rowe, F. Harmon, C.L. Cass, J.C. Sedbrook, M.E. Byrne, S. Walsh, J. Higgins, M. Bevan, P. Li, T. Brutnell, T. Unver, H. Budak, H. Belcram, M. Charles, B. Chalhoub, I. Baxter, The International Brachypodium Initiative, Genome sequencing and analysis of the model grass *Brachypodium distachyon*, *Nature*. 463 (2010) 763–768, <https://doi.org/10.1038/nature08747>.
- [321] T. Wan, Z.-M. Liu, L.-F. Li, A.R. Leitch, J.L. Leitch, R. Lohaus, Z.-J. Liu, H.-P. Xin, Y.-B. Gong, Y. Liu, W.-C. Wang, L.-Y. Chen, Y. Yang, L.J. Kelly, J. Yang, J.-L. Huang, Z. Li, P. Liu, L. Zhang, H.-M. Liu, H. Wang, S.-H. Deng, M. Liu, J. Li, L. Ma, Y. Liu, Y. Lei, W. Xu, L.-Q. Wu, F. Liu, Q. Ma, X.-R. Yu, Z. Jiang, G.-Q. Zhang, S.-H. Li, R.-Q. Li, S.-Z. Zhang, Q.-F. Wang, Y. Van de Peer, J.-B. Zhang, X.-M. Wang, A genome for gnetophytes and early evolution of seed plants, *Nat. Plants* 4 (2018) 82–89, <https://doi.org/10.1038/s41477-017-0097-2>.
- [322] A.Z. Worden, J.-H. Lee, T. Mock, P. Rouzé, M.P. Simmons, A.L. Aerts, A.E. Allen, M.L. Cuvelier, E. Derelle, M.V. Everett, E. Foulon, J. Grimwood, H. Gundlach, B. Henrissat, C. Napoli, S.M. McDonald, M.S. Parker, S. Rombauts, A. Salamov, P. V. Dassow, J.H. Badger, P.M. Coutinho, E. Demir, I. Dubchak, C. Gentemann, W. Eikrem, J.E. Gready, U. John, W. Lanier, E.A. Lindquist, S. Lucas, K.F. X. Mayer, H. Moreau, F. Not, R. Otitlar, O. Panaud, J. Pangilinan, I. Paulsen, B. Piegu, A. Poliakov, S. Robbens, J. Schmutz, E. Toulza, T. Wyss, A. Zelensky, K. Zhou, E.V. Armbrust, D. Bhattacharya, U.W. Goodenough, Y.V. de Peer, I. V. Grigoriev, Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes micromonas, *Science* 324 (2009) 268–272, <https://doi.org/10.1126/science.1167222>.
- [323] X. Argout, J. Salse, J.-M. Aury, M.J. Guiltinan, G. Droc, J. Gouzy, M. Allegre, C. Chaparro, T. Legavre, M.N. Maximova, M. Abrouk, F. Murat, O. Fouet, J. Poulain, M. Ruiz, Y. Roguet, M. Rodier-Goud, J.F. Barbosa-Neto, F. Sabot, D. Kudrna, J.S.S. Ammiraju, S.C. Schuster, J.E. Carlson, E. Sallet, T. Schiex, A. Dievart, M. Kramer, L. Gelley, Z. Shi, A. Bérard, C. Viot, M. Boccara, A. M. Risterucci, V. Guignon, X. Sabau, M.J. Axtell, Z. Ma, Y. Zhang, S. Brown, M. Bourge, W. Golser, X. Song, D. Clement, R. Rivallan, M. Tahj, J.M. Akaza, B. Pitollat, K. Gramacho, A. D'Hont, D. Brunel, D. Infante, I. Kebe, P. Costet, R. Wing, W.R. McCombie, E. Guiderdoni, F. Quétier, O. Panaud, P. Wincker, S. Bocs, C. Lanaud, The genome of *Theobroma cacao*, *Nat. Genet.* 43 (2011) 101–108, <https://doi.org/10.1038/ng.736>.
- [324] G. Blanc, G. Duncan, I. Agarkova, M. Borodovsky, J. Gurnon, A. Kuo, E. Lindquist, S. Lucas, J. Pangilinan, J. Polle, A. Salamov, A. Terry, T. Yamada, D.D. Dunigan, I. V. Grigoriev, J.-M. Claverie, J.L. Van Etten, The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex, *Plant Cell* 22 (2010) 2943–2955, <https://doi.org/10.1105/tpc.110.076406>.
- [325] T. Dagan, M. Roettger, K. Stucken, G. Landan, R. Koch, P. Major, S.B. Gould, V. V. Goremykin, R. Rippka, N. Tandeau de Marsac, M. Gugger, P.J. Lockhart, J. F. Allen, I. Brune, I. Maus, A. Puhler, W.F. Martin, Genomes of stigonematalean Cyanobacteria (Subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids, *Genome Biol. Evol.* 5 (2013) 31–44, <https://doi.org/10.1093/gbe/evs117>.
- [326] J.C. Dohm, A.E. Minoche, D. Holtgräwe, S. Capella-Gutiérrez, F. Zkrzewski, H. Tafer, O. Rupp, T.R. Sörensen, R. Stracke, R. Reinhardt, A. Goesmann, T. Kraft, B. Schulz, P.F. Stadler, T. Schmidt, T. Gabaldón, H. Lehrach, B. Weisshaar, H. Himmelbauer, The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*), *Nature* 505 (2014) 546–549, <https://doi.org/10.1038/nature12817>.
- [327] J. Kreplak, M.-A. Madoui, P. Cápál, P. Novák, K. Labadie, G. Aubert, P.E. Bayer, K. K. Gali, R.A. Syme, D. Main, A. Klein, A. Bérard, I. Vrbová, C. Fournier, L. d'Agata, C. Belser, W. Berrabah, H. Toegelová, Z. Milec, J. Vrána, H. Lee, A. Kougbeadjio, M. Térézol, C. Huneau, C.J. Turo, N. Mohellibi, P. Neumann, M. Falque, K. Gallardo, R. McGee, B. Tar'an, A. Bendahmane, J.-M. Aury, J. Batley, M.-C. Le Paslier, N. Ellis, T.D. Warkentin, C.J. Coyne, J. Salse, D. Edwards, J. Lichtenzweig, J. Macas, J. Dolezel, P. Wincker, J. Burstin, A reference genome for pea provides insight into legume genome evolution, *Nat. Genet.* 51 (2019) 1411–1422, <https://doi.org/10.1038/s41588-019-0480-1>.
- [328] H. Li, F. Jiang, P. Wu, K. Wang, Y. Cao, A high-quality genome sequence of model legume *Lotus japonicus* (MG-20) provides insights into the evolution of root nodule symbiosis, *Genes* 11 (2020) 483, <https://doi.org/10.3390/genes11050483>.
- [329] J.C. Meeks, J. Elhai, T. Thiel, M. Potts, F. Larimer, J. Lamerdin, P. Predki, R. Atlas, An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium, *Photosynth. Res* 70 (2001) 85–106, <https://doi.org/10.1023/A:1013840025518>.
- [330] H. Moreau, B. Verhelst, A. Couloux, E. Derelle, S. Rombauts, N. Grimsley, M. V. Bel, J. Poulain, M. Katinka, M.F. Hohmann-Mariotti, G. Piganeau, P. Rouzé, C. D. Silva, P. Wincker, Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage, *Genome Biol.* 13 (2012) R74, <https://doi.org/10.1186/gb-2012-13-8-r74>.
- [331] R.I. Ponce-Toledo, P. Deschamps, P. Lopez-Garcia, Y. Zivanovic, K. Benzerara, D. Moreira, An early-branching freshwater cyanobacterium at the origin of plastids, *Current Biol* 27 (2017) 386–391, <https://doi.org/10.1016/j.cub.2016.11.056>.
- [332] S.E. Prochnik, J. Umen, A.M. Nedelcu, A. Hallmann, S.M. Miller, I. Nishii, P. Ferris, A. Kuo, T. Mitros, L.K. Fritz-Laylin, U. Hellsten, J. Chapman, O. Silva, P. Rensing, A. Terry, J. Pangilinan, V. Kapitonov, J. Jurka, A. Salamov, H. Shapiro, J. Schmutz, J. Grimwood, E. Lindquist, S. Lucas, I. V. Grigoriev, R. Schmitt, D. Kirk, D.S. Rokhsar, Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*, *Science* 329 (2010) 223–226, <https://doi.org/10.1126/science.1188800>.
- [333] L. Ran, J. Larsson, T. Vigil-Stenman, J.A.A. Nylander, K. Ininbergs, W.-W. Zheng, A. Lapidus, S. Lowry, R. Haselkorn, B. Bergman, Genome erosion in a nitrogen-fixing vertically transmitted endosymbiotic multicellular cyanobacterium, *PLoS ONE* 5 (2010), e11486, <https://doi.org/10.1371/journal.pone.0011486.s005>.
- [334] N. Sierro, J.N.D. Battey, S. Ouadi, N. Bakaher, L. Bovet, A. Willig, G. Goepfert, M. C. Peitsch, N.V. Ivanov, The tobacco genome sequence and its comparison with those of tomato and potato, *Nat. Commun.* 5 (2014) 3833, <https://doi.org/10.1038/ncomms4833>.
- [335] T. Slotte, K.M. Hazzouri, J.A. Ågren, D. Koenig, F. Maumus, Y.-L. Guo, K. Steige, A.E. Platts, J.S. Escobar, L.K. Newman, W. Wang, T. Mandáková, E. Vello, L. M. Smith, S.R. Henz, J. Steffen, S. Takuno, Y. Brandvain, G. Coop, P. Andolfatto, T.T. Hu, M. Blanchette, R.M. Clark, H. Quesneville, M. Nordborg, B.S. Gaut, M. A. Lysak, J. Jenkins, J. Grimwood, J. Chapman, S. Prochnik, S. Shu, D. Rokhsar, J. Schmutz, D. Weigel, S.I. Wright, The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution, *Nat. Genet.* 45 (2013) 831–835, <https://doi.org/10.1038/ng.2669>.
- [336] The Brassica rapa Genome Sequencing Project Consortium, The genome of the mesopolyploid crop species *Brassica rapa*, *Nat. Genet.* 43 (2011) 1035–1039, <https://doi.org/10.1038/ng.919>.
- [337] N.D. Young, F. DeBelle, G.E.D. Oldroyd, R. Geurts, S.B. Cannon, M.K. Udvardi, V. A. Benedito, K.F.X. Mayer, J. Gouzy, J. Schoof, Y. Van de Peer, S. Proost, D. R. Cook, B.C. Meyers, M. Spannagl, F. Cheung, S. De Mita, V. Krishnakumar, H. Gundlach, S. Zhou, J. Mudge, A.K. Bharti, J.D. Murray, M.A. Naoumkina, B. Rosen, K.A.T. Silverstein, H. Tang, S. Rombauts, P.X. Zhao, P. Zhou, V. Barbe, P. Bardou, M. Bechner, A. Bellec, A. Berger, H. Bergès, S. Bidwell, T. Bisseling, N. Choise, A. Couloux, R. Denny, S. Deshpande, X. Dai, J.J. Doyle, A.-M. Dudez, A.D. Farmer, S. Fouteau, C. Franken, C. Gibelin, J. Gish, S. Goldstein, A. J. González, P.J. Green, A. Hallab, M. Hartog, A. Hua, S.J. Humphray, D.-H. Jeong, Y. Jing, A. Jöcker, S.M. Kenton, D.-J. Kim, K. Klee, H. Lai, C. Lang, S. Lin, S.L. Macmill, G. Magdelenat, L. Matthews, J. McCorrison, E.L. Monaghan, J.-H. Mun, F.Z. Najjar, C. Nicholson, C. Noirot, M. O'Benness, C.R. Paule, J. Poulain, F. Prion, B. Qin, C. Qu, E.F. Retzel, C. Riddle, E. Sallet, S. Samain, N. Samson, I. Sanders, O. Saurat, C. Scarpelli, T. Schiex, B. Segurens, A.J. Severin, D.J. Sherrier, R. Shi, S. Sims, S.R. Singer, S. Sinharoy, L. Sterck, A. Viollet, B.-B. Wang, K. Wang, M. Wang, X. Wang, J. Warfsmann, J. Weissenbach, D.D. White, J.D. White, G.B. Wiley, P. Wincker, Y. Xing, L. Yang, Z. Yao, F. Ying, J. Zhai, L. Zhou, A. Zuber, J. Dénarié, R.A. Dixon, G.D. May, D.C. Schwartz, J. Rogers, F. Quétier, C.D. Town, B.A. Roe, The *Medicago* genome provides insight into the evolution of rhizobial symbioses, *Nature* 480 (2011) 520–524, <https://doi.org/10.1038/nature10625>.
- [338] Smýkalová, I., Ludvíková, M., Ondráčková, E., Klejduš, B., Bonhomme, S., Kronusová, O., Soukup, A., Rozmoš, M., Guzow-Krzemińska, B., Matúšová, R., 2017. Green microalga *Trebouxia* sp. produces strigolactone-related compounds (preprint). bioRxiv (2017) (<https://doi.org/10.1101/195883>).
- [339] K.M. Davies, R. Jibrán, Y. Zhou, N.W. Albert, D.A. Brummell, B.R. Jordan, J. L. Bowman, K.E. Schwinn, K.E. The evolution of flavonoid biosynthesis: a bryophyte perspective, *Front. Plant Sci.* 11 (2020) 7, <https://doi.org/10.3389/fpls.2020.00007>.
- [340] G. Soriano, M.-Á. Del-Castillo-Alonso, L. Monforte, E. Núñez-Olivera, J. Martínez-Abaigar, First data on the effects of ultraviolet radiation on phenolic compounds in the model hornwort *Anthoceros agrestis*, *Cryptogam., Bryol.* 39 (2018) 201–211, <https://doi.org/10.7872/cryb/v39.iss2.2018.201>.
- [341] J.-K. Weng, J.P. Noel, Chemodiversity in *Selaginella*: a reference system for parallel and convergent metabolic evolution in terrestrial plants, *Front. Plant Sci.* 4 (2013) 119, <https://doi.org/10.3389/fpls.2013.00119>.
- [342] E. Güngör, P. Brouwer, L.W. Dijkhuizen, D.C. Shaffar, K.G.J. Nierop, R.C.H. Vos, J. Sastre Torano, I.M. Meer, H. Schlupemann, *Azolla* ferns testify: seed plants and ferns share a common ancestor for leucoanthonyaninid reductase enzymes, *New Phytol.* 229 (2021) 1118–1132, <https://doi.org/10.1111/nph.16896>.

- [343] T. Iwashina, The structure and distribution of the flavonoids in plants, *J. Plant Res* 113 (2000) 287–299, <https://doi.org/10.1007/PL00013940>.
- [344] M. Jung, H.D. Zinsmeister, H. Geiger, New three- and tetraoxygenated coumarin glucosides from the mosses *Atrichum undulatum* and *Polytrichum formosum*, *Z. für Naturforsch. C.* 49 (1994) 697–702, <https://doi.org/10.1515/znc-1994-11-1201>.
- [345] D. Menzel, R. Kazlauskas, J. Reichelt, Coumarins in the siphonalean green algal family Dasycladaceae Kützinger (Chlorophyceae), *Bot. Mar.* 26 (1983) 23–29, <https://doi.org/10.1515/botm.1983.26.1.23>.
- [346] E. Perez-Rodriguez, J. Aguilera, I. Gómez, F.L. Figueroa, Excretion of coumarins by the Mediterranean green alga *Dasycladus vermicularis* in response to environmental stress, *Mar. Biol.* 139 (2001) 633–639, <https://doi.org/10.1007/s002270100588>.
- [347] E. Perez-Rodriguez, J. Aguilera, F.L. Figueroa, Tissue localization of coumarins in the green alga *Dasycladus vermicularis* (Scopoli) Krasser: a photoprotective role?, 2003, *J. Exp. Bot.* 54 (2003) 1093–1100, <https://doi.org/10.1093/jxb/erg111>.
- [348] H. Lijun, The auxin concentration in sixteen Chinese marine algae, *Chin. J. Oceanol. Limnol.* 24 (2006) 329–332, <https://doi.org/10.1007/bf02842637>.
- [349] H. Mazur-Marzec, A. Konop, R. Synak, Indole-3-acetic acid in the culture medium of two axenic green microalgae, *J. Appl. Phycol.* 13 (2001) 35–42, <https://doi.org/10.1023/A:1008199409953>.
- [350] G. Pichler, W. Stögl, F. Candotto Carniel, L. Muggia, C.G. Ametrano, A. Holzinger, M. Tretiach, I. Kranner, Abundance and extracellular release of phytohormones in aero-terrestrial microalgae (Trebouxiophyceae, Chlorophyta) as a potential chemical signaling source, *J. Phycol.* 56 (2020) 1295–1307, <https://doi.org/10.1111/jpy.13032>.
- [351] S.R. Cutler, P.L. Rodriguez, R.R. Finkelstein, S.R. Abrams, Abscisic Acid: Emergence of a Core Signaling Network, *Annu. Rev. Plant Biol.* 61 (2010) 651–679, <https://doi.org/10.1146/annurev-arplant-042809-112122>.
- [352] C. Lind, I. Dreyer, E.J. López-Sanjurjo, K. von Meyer, K. Ishizaki, T. Kohchi, D. Lang, Y. Zhao, I. Kreuzer, K.A.S. Al-Rasheid, H. Ronne, R. Reski, J.-K. Zhu, D. Geiger, R. Hedrich, Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure, *Curr. Biol.* 25 (2015) 928–935, <https://doi.org/10.1016/j.cub.2015.01.067>.
- [353] A. Shinozawa, R. Otake, D. Takezawa, T. Umezawa, K. Komatsu, K. Tanaka, A. Amagai, S. Ishikawa, Y. Hara, Y. Kamisugi, A.C. Cuming, K. Hori, H. Ohta, F. Takahashi, K. Shinozaki, T. Hayashi, T. Taji, Y. Sakata, SnRK2 protein kinases represent an ancient system in plants for adaptation to a terrestrial environment, *Commun. Biol.* (2019) 1–13, <https://doi.org/10.1038/s42003-019-0281-1>.

4.2 **Publication V:** The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families

This research paper was published in the Journal “The Plant Journal” in June 2021. The full article as well as all supplementary figures and supplementary datasets can be found online:

<https://doi.org/10.1111/tpj.15387>

Contribution of Janine Fürst-Jansen, shared first author

J. M. R. Fürst-Jansen processed phylogenies and designed the layout of all main figures that show a phylogenetic analysis (figures 2, 3, 4, 5, 6, 7, 8, 9, as well as supplementary figures S2, S8). She wrote parts of the manuscript and critically read and revised the manuscript.

The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families

Sophie de Vries^{1,2,†} , Janine M. R. Fürst-Jansen^{2,†} , Iker Irisarri^{2,3,†} , Amra Dhabalia Ashok² , Till Ischebeck^{4,5,6} , Kirstin Feussner^{4,5} , Ilka N Abreu⁴ , Maike Petersen⁷ , Ivo Feussner^{4,5,6}  and Jan de Vries^{2,3,8*} 

¹Population Genetics, Heinrich-Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany,

²Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Goettingen, Goldschmidtstr. 1, 37077 Goettingen, Germany,

³University of Goettingen, Campus Institute Data Science (CIDAS), Goldschmidtstr. 1, 37077 Goettingen, Germany,

⁴Department of Plant Biochemistry, University of Goettingen, Albrecht-von-Haller-Institute for Plant Sciences, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany,

⁵Goettingen Center for Molecular Biosciences (GZMB), Goettingen Metabolomics and Lipidomics Laboratory, University of Goettingen, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany,

⁶Department of Plant Biochemistry, Goettingen Center for Molecular Biosciences (GZMB), University of Goettingen, Justus-von-Liebig Weg 11, 37077 Goettingen, Germany,

⁷Institut für Pharmazeutische Biologie und Biotechnologie, Philipps-Universität Marburg, Robert-Koch-Str. 4, 35037 Marburg, Germany, and

⁸Department of Applied Bioinformatics, Goettingen Center for Molecular Biosciences (GZMB), University of Goettingen, Goldschmidtstr. 1, 37077 Goettingen, Germany

Received 26 January 2021; revised 11 June 2021; accepted 21 June 2021; published online 24 June 2021.

*For correspondence (e-mail devries.jan@uni-goettingen.de).

†These authors contributed equally.

SUMMARY

Land plants constantly respond to fluctuations in their environment. Part of their response is the production of a diverse repertoire of specialized metabolites. One of the foremost sources for metabolites relevant to environmental responses is the phenylpropanoid pathway, which was long thought to be a land-plant-specific adaptation shaped by selective forces in the terrestrial habitat. Recent data have, however, revealed that streptophyte algae, the algal relatives of land plants, have candidates for the genetic toolkit for phenylpropanoid biosynthesis and produce phenylpropanoid-derived metabolites. Using phylogenetic and sequence analyses, we here show that the enzyme families that orchestrate pivotal steps in phenylpropanoid biosynthesis have independently undergone pronounced radiations and divergence in multiple lineages of major groups of land plants; sister to many of these radiated gene families are streptophyte algal candidates for these enzymes. These radiations suggest a high evolutionary versatility in the enzyme families involved in the phenylpropanoid-derived metabolism across embryophytes. We suggest that this versatility likely translates into functional divergence, and may explain the key to one of the defining traits of embryophytes: a rich specialized metabolism.

Keywords: phenylpropanoid biosynthesis, plant evolution, evolution of gene families, evo-physio.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit <https://doi.org/10.1111/tpj.15423>

INTRODUCTION

A diverse profile of specialized metabolites is one of the characteristics of land plants (embryophytes). Almost any aspect of the biology of land plants is underpinned by specialized metabolites—be it the phytohormones that are major modulators upstream in various genetic hierarchies (Berens

et al., 2017; Blázquez *et al.*, 2020; Scheres and van der Putten, 2017) or pigments that give land plants their color and attune photochemical properties (Jahns and Holzwarth, 2012).

A key aspect of the biological relevance of most specialized metabolites is their use under challenging environmental conditions. Indeed, the elaboration of their specialized metabolism is considered one of the drivers for

the substantial radiation of embryophytes on land (Weng, 2013). Moreover, a diversity of specialized metabolism likely played a key role during the earliest steps of plants on land—allowing for the production of compounds that protected land plants against the challenges of the terrestrial environment such as drought and increased UV radiation (Fürst-Jansen *et al.*, 2020; Jiao *et al.*, 2020; Rensing, 2018; de Vries and Archibald, 2018). A major pathway giving rise to a variety of specialized metabolites that act in warding off environmental stressors is the biosynthesis of phenylpropanoids (Dixon *et al.*, 2002; Dixon and Paiva, 1995; Vogt, 2010).

The phenylpropanoid pathway is the source of precursors for thousands of metabolites with multifaceted functions, and accounts for about 40% of organic carbon on earth (Vogt, 2010). One facet of these functions is that phenylpropanoid-derived compounds act as structural polymers, foremost among which are the different types of lignin (Ralph *et al.*, 2004; Vanholme *et al.*, 2012, 2019). Another prominent facet is that these metabolites act as UV-protecting substances. While some of the best-known UV screens are flavonoids, various other compounds stemming from the phenylpropanoid pathway are equally potent UV protectants (Booij-James *et al.*, 2000; Sheahan, 1996; Sytar *et al.*, 2018; Xue *et al.*, 2020). The list of links between phenylpropanoid-derived compounds and the response to environmental challenges could be continued; in fact, the response to almost any abiotic stressor that plants face in the terrestrial habitat involves the action of phenylpropanoid-derived compounds (for comprehensive reviews, see Dixon and Paiva, 1995; Vogt, 2010). Furthermore, phenylpropanoid-derived compounds are involved in the defense responses against plant pathogens in many land plant lineages (Carella *et al.*, 2019; Danielsson *et al.*, 2011; König *et al.*, 2014; Overdijk *et al.*, 2016; Ponce De León *et al.*, 2012).

All embryophytes make use of the enzymatic routes in the phenylpropanoid pathway. For example, the utilization of flavonoids under UV stress appears to be a conserved response across Embryophyta (Clayton *et al.*, 2018; Wolf *et al.*, 2010). However, not all embryophytes produce the same compounds under the same stress conditions—on the contrary, the diversity of compounds is immense. Major differences in the biosynthesis of phenylpropanoid-derived compounds occur in distinct lineages of land plants. This includes specialized roles such as the flower coloration determining anthocyanins that attract pollinators (Miller *et al.*, 2011; Sheehan *et al.*, 2012); such a role of anthocyanins is obviously limited to flowering plants and can vary even among closely related species (Saito and Harborne, 1992). That said, Piatkowski *et al.* (2020) phylogenetically inferred that orthologs for the entire anthocyanin biosynthesis pathway were already present in the ancestor of seed plants, and more than half of the

important orthogroups were already present in the most recent common ancestor of all land plants. An important recent insight into the deep evolutionary roots of flavonoid biosynthesis was the discovery of auronidins—a class of red flavonoid pigments that are synthesized in the bryophyte *Marchantia polymorpha* (Berland *et al.*, 2019). Further, for example, Renault *et al.* (2017a) reported on the enrichment of the *Physcomitrium patens* (moss) cuticle in phenolic compounds—an enrichment that hinges on the action of a cytochrome P450 enzyme that is orthologous to enzymes that act in lignin biosynthesis; the production of lignin might trace its evolutionary roots back to an ancient set of enzymes acting in the production of complex, phenol-enriched polymers (Renault *et al.*, 2019). Carella *et al.* (2019) showed that the liverwort model plant *M. polymorpha* triggers phenylpropanoid biosynthesis upon attack by the oomycete phytopathogen *Phytophthora palmivora*. Similar responses towards phytopathogens are known from gymnosperms (Oliva *et al.*, 2015) and angiosperms (Bednarek *et al.*, 2005; Carella *et al.*, 2019; Chezem *et al.*, 2017; Dixon and Paiva, 1995; Kaur *et al.*, 2010). Thus, all land plants use the core framework of the phenylpropanoid pathway to produce—often lineage-specific—variations of phenylpropanoid derivatives that aid in response to biotic and abiotic stressors.

The production of the chemical repertoire of land plants is often catalyzed by members of large enzyme-coding gene families (Nelson and Werck-Reichhart, 2011; Renault *et al.*, 2017b; Shockey *et al.*, 2003), and this also seems to be the case for the enzymes involved in phenylpropanoid biosynthesis (Hamberger *et al.*, 2007; Vogt, 2010; Xu *et al.*, 2009). It is thus conceivable that various adaptive forces have shaped the families of enzymes that act in the phenylpropanoid pathway, leading to multiple independent cases of sub- and neofunctionalization (Rensing, 2014). An inference of the common (minimal) set of enzymes that were present in the last common ancestors (LCAs) of: (i) streptophytes; (ii) land plants and their closest streptophyte algal relatives; and (iii) land plants can thus shed light on which enzymatic building blocks evolution acted upon to give rise to the elaborate chassis of the phenylpropanoid pathway.

The phenylpropanoid pathway has long been considered to be specific to Embryophyta. However, homologs of the genes coding for the enzymes that constitute the embryophytic phenylpropanoid pathway can be found in extant algal relatives of land plants, suggesting that they were already present in a common ancestor shared by streptophyte algae and land plants (Maeda and Fernie, 2021; Renault *et al.*, 2019; de Vries *et al.*, 2017). Since the beginning of 2020, we have genome data from all major lineages of Streptophyta—except for Coleochaetophyceae (Szövényi *et al.*, 2021); only using this extended repertoire of species and sequences allows us to pinpoint which sub-families and/or which ancestral enzymes of multiple

subfamilies were present in the aforementioned LCAs. Compounds that, in land plants, emerge from the phenylpropanoid pathway are indeed found in algae; these include flavonoids and lignin-like compounds in streptophyte algae (Delwiche *et al.*, 1989; Jiao *et al.*, 2020; Sørensen *et al.*, 2011) and core phenylpropanoid building blocks as well as flavonoids in a phylo diverse set of algae (Goiris *et al.*, 2014). Interestingly, lignin-like compounds were even found in distantly-related red macroalgae (Martone *et al.*, 2009)—although this is likely a case of convergence that builds on an unknown enzymatic framework. However, even within the green lineage (Chloroplastida), the question of the deep evolutionary roots of the phenylpropanoid pathway is wide open.

Investigations of the algal relatives of land plants have strongly benefitted from recent progress in phylogenomics on plants and algae. A major outcome of these recent phylogenomic analyses was that the Zygnematophyceae have been pinpointed as the class of algae most closely related to land plants (Leebens-Mack *et al.*, 2019; Wickett *et al.*, 2014; Wodniok *et al.*, 2011). Hand in hand with these phylogenomic efforts went the generation of genomic (Cheng *et al.*, 2019; Hori *et al.*, 2014; Jiao *et al.*, 2020; Nishiyama *et al.*, 2018; Wang *et al.*, 2020) and transcriptomic data on streptophyte algae (Ju *et al.*, 2015; Rippin *et al.*, 2017, 2019; de Vries *et al.*, 2018, 2020). Additionally, critical gaps in the land plant tree of life have been filled; this includes recent publications of genomes of liverworts (Bowman *et al.*, 2017), ferns (Li *et al.*, 2018) and hornworts (Li *et al.*, 2020; Szövényi *et al.*, 2015; Zhang *et al.*, 2020). These data allow for the fine-grained tracing of the evolution of key plant enzyme families across the green tree of life. Recent studies have illuminated the diversity of enzymes in the routes towards flavonoids and anthocyanins as well as the PAL-dependent pathway of salicylic acid biosynthesis via benzoic acid (Güngör *et al.*, 2021; Piatkowski *et al.*, 2020; de Vries *et al.*, 2021).

In this study, we infer the evolutionary history of 11 critical enzyme families known to be woven into the mesh of routes from phenylpropanoids to lignin biosynthesis in land plants; we have paid particular attention to the routes leading to the biosynthesis of lignin. We use the expanded diversity of genomic and transcriptomic data from land plants as well as streptophyte and chlorophyte algae to infer the origin of these large gene families. The datasets were chosen in a manner that they cover the breadth of streptophyte diversity while providing a balanced sampling; the latter is especially relevant in light of the high number of genomes available for flowering plants. We aimed to include at least one representative of each of the major lineages of streptophytes in the datasets we surveyed. Our data pinpoint deep homologs of candidate enzymes in streptophyte algae for L-phenylalanine ammonia-lyase (PAL), 4-coumarate-CoA ligase (4CL), caffeoyl-CoA O-methyltransferase (CCoAOMT); further, for streptophyte and chlorophyte algae, we

pinpoint homologs for cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD) and potentially relevant monoacylglycerol lipases (MAGLs). Further, we find that often the functionally characterized enzymes of the core phenylpropanoid and lignin biosynthesis routes derive from lineage-specific radiations, limiting the inference of function outside the model system. Nonetheless we could infer which subfamilies were present in the LCAs along the trajectory of streptophyte evolution, even though ancestral functional inference was limited. That said, for enzyme families with deep homologs, we approximated the function through domain prediction and the conservation (or lack thereof) of key residues of known functional importance. We found that all enzyme families underwent several lineage-specific expansions and losses as well as bursts in growth of enzyme families that occurred early during the radiation of land plants. We hypothesize that lineage-specific expansions in these enzyme families are linked with the diversity of lineage-specific phenylpropanoid derivatives and functions that occur in the species analyzed here.

RESULTS AND DISCUSSION

The checkered occurrence of PAL among streptophyte algae

The conversion of the aromatic amino acid phenylalanine and/or tyrosine into cinnamate and/or *p*-coumarate is the first step of the plant phenylpropanoid pathway (Figure 1). This first committed step is catalyzed by PAL and the bifunctional L-phenylalanine/L-tyrosine ammonia-lyase (PTAL; Barros and Dixon, 2020). For a long time, it was thought that among Chloroplastida, PAL/PTAL were limited to land plants; their gain was considered to have occurred via a lateral gene transfer event that had happened at the base of the land plant clade (Emiliani *et al.*, 2009). Recently, however, genes coding for putative PAL-like enzymes were detected in streptophyte algae, such as the filamentous streptophyte alga *Klebsormidium nitens* (de Vries *et al.*, 2017). In light of the recent surge in available genomes from across the green tree of life, we set out to explore the evolutionary history of PAL.

Using AtPAL1 as a bait sequence, we screened protein data from diverse land plants and all streptophyte algal genomes available via BLAST. Hits among streptophyte algae fell into three categories: (i) proteins of between 480 (*Klebsormidium nitens* PAL, kfl00104_0290_v1.1) and 527 amino acids (two homologs in *Chara braunii*; g57646_t1 and g34530_t1); (ii) short proteins such as ME1156409C09523 of *Mesotaenium endlicherianum*, which is 184 amino acids in length; (iii) long proteins of between 991 (*Chlorokybus atmophyticus* Chrsp482S06115) and 1115 amino acids (*K. nitens* kfl00024_0250_v1.1). Proteins falling into the third category are fusions of an aromatic amino acid lyase

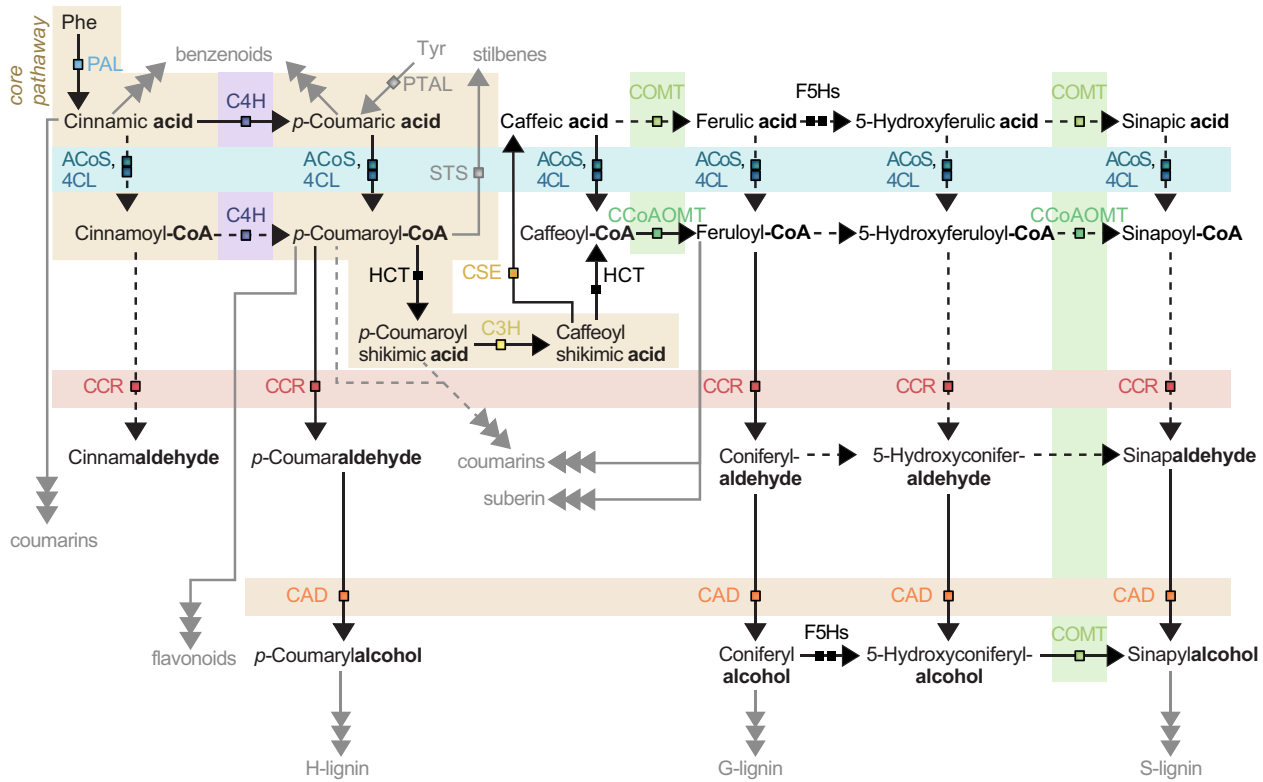


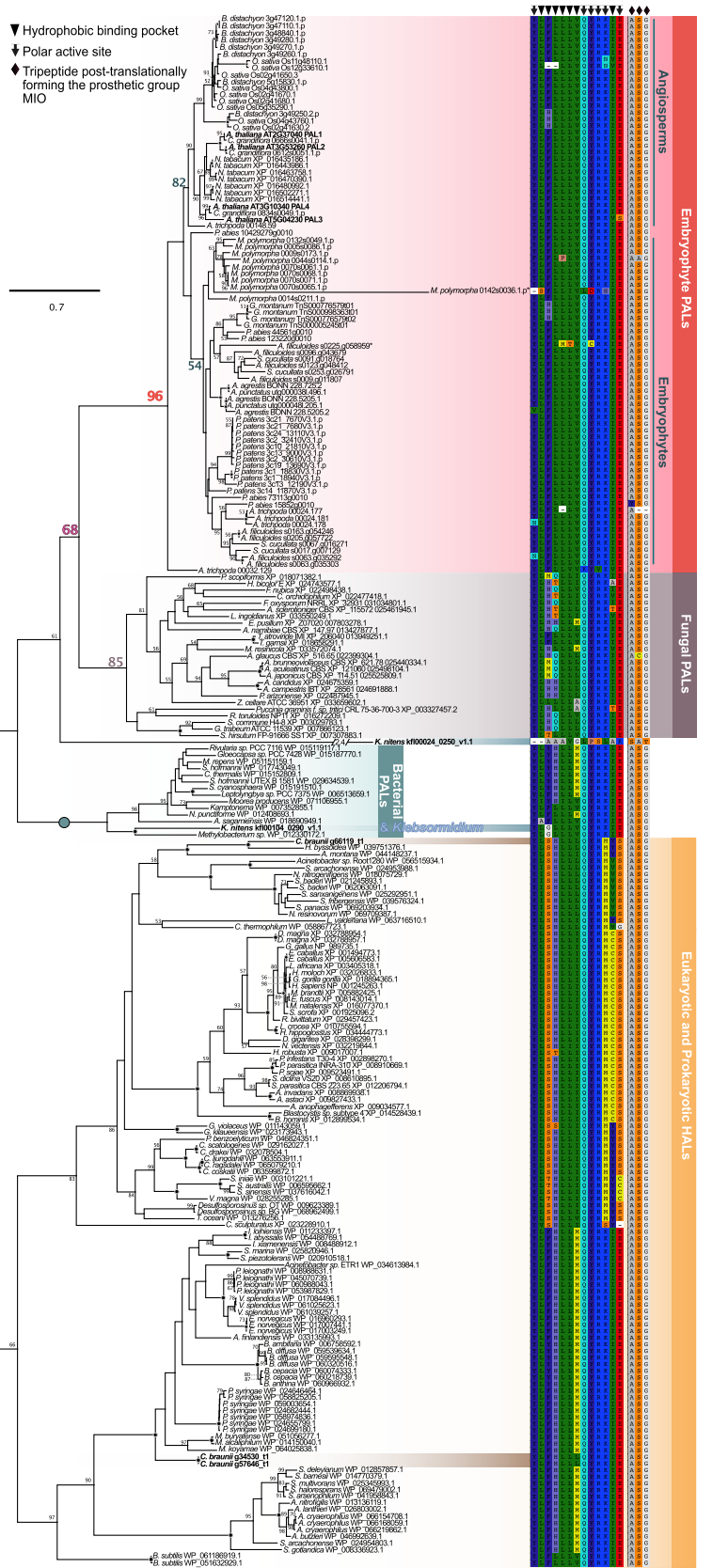
Figure 1. Enzymes involved in the biosynthesis of phenylpropanoid-derived compounds investigated here. A simplified schematic view of the phenylpropanoid pathway and its routes to different derivatives are shown. Boxes indicate enzyme families, which are mentioned above each box and color-coded. Their coloration is the same as in Figure 10. Dotted lines indicate putative/ambiguous steps in the pathway. As highlighted by (i) the alignment of similar steps and (ii) the background shadings, the same/similar enzymes act at different steps in the pathway; this speaks to the modularity, versatility and recurring themes in the pathway.

domain and a putative tRNA synthetase; homologs of such also occur in land plants (e.g. AT3G02760 and Os05g0150900). The short proteins of category ‘(iii)’ are found in some Zygnematophyceae and are proteins of unknown function with a putative HAL domain. Genomic artefacts leading to this result can be excluded given that these types of protein-encoding genes have been recovered for independent Zygnematophyceae that are likely >500 million years divergent from one another. Interestingly, despite its short size, querying the protein ME1156409C09523 from *M. endlicherianum* in I-TASSER (Iterative Threading ASSEMBLY Refinement; Zhang, 2008) recovered *Pseudomonas putida* HAL with a good TM-score (0.865; 1GKM; Baedeker and Schulz, 2002). Some of the residues that have been highlighted by Jun *et al.* (2018) and Nagy *et al.* (2019) for being active site residues (residues Y9, L35, S38, H39, K150, I154) have also been recovered in this short PAL-like sequence, and are predicted to fall into a hypothetical catalytic pocket region according to I-TASSER (Figure S1a). Yet, the tripeptide ASG that forms the post-translational prosthetic group 3,5-dihydro-5-methylene-4H-imidazol-4-one (MIO) essential for the catalytic activity of PAL and HAL (Schwede *et al.*, 1999) is not present in the short protein of *M. endlicherianum*.

The proteins in category ‘(i)’ are those with the highest identity to *bona fide* land plant PALs. This category includes the promising PAL candidate kfl00104_0290_v1.1 (de Vries *et al.*, 2017), which recovered PAL from *Petroselinum crispum* as its closest structural analog (1W27; Ritter and Schulz, 2004; TM-score: 0.927) in I-TASSER (Figure S1b). All residues involved in forming the binding pocket are present, map to the predicted binding pocket and are conserved in kfl00104_0290_v1.1 (Figures S1b and 2), with the exception of one Phe that is substituted by Gly (position 114) in *K. nitens*. This particular residue is important for determining tyrosine aminomutase, HAL and fungal PALs from plant PALs (Jun *et al.*, 2018; de Vries *et al.*, 2021). It is noteworthy that kfl00104_0290_v1.1 has with Tyr113 another aromatic amino acid adjacent to G114. The C-terminus of kfl00104_0290_v1.1 also includes the potentially relevant residues for stabilizing the carboxy group (R323) and the counter ions to the E456 (hydrogen bonding), Q317 and K427 (Figure S1b), allowing for the possibility that kfl00104_0290_v1.1 might add key residues to the active site assuming that this PAL homolog forms a homotetramer similar to canonical PALs. Moreover, the catalytical MIO group is also recovered as an integral part

Figure 2. A phylogenetic framework for the origin of streptophyte phenylalanine ammonia-lyase (PAL).

PAL and histidine ammonia lyase (HAL) homologs were sampled from 15 land plants, seven streptophyte algae and eight chlorophytes. Among Chloroplastida, PAL homologs were only recovered from genomes of land plants and the streptophyte algae *Klebsormidium nitens* and *Chara braunii*. From all detected homologs, a maximum likelihood phylogeny was computed using LG+I+G4 as model for protein evolution (chosen according to BIC); the phylogenetic tree was rooted with HAL. One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; the clade of *bona fide* PALs was assigned the color red; *Klebsormidium* and *Chara* sequences are highlighted with a teal and brown background, respectively. On the right we show key residues for substrate binding and function as reported by Jun *et al.* (2018) and Nagy *et al.* (2019). Additionally, we highlight the tripeptide Ala, Ser, Gly from which, post-translationally, the prosthetic electrophilic group 4-methylideneimidazole-5-one (MIO; Schwede *et al.*, 1999) is formed; sequences labeled with an asterisk have a duplication in amino acids that form the tripeptide.



of the binding pocket for the PAL candidate of *K. nitens* (Figures 2 and S1b). We therefore set out to further explore these PAL-like candidates, which noteworthy were only found in the genome of *K. nitens* and *C. braunii* and no other streptophyte algal genome.

To understand the evolutionary history of PAL in streptophytes, we computed a maximum likelihood phylogeny (Figure 2). The phylogenetic analysis included the aforementioned PAL homologs from diverse Streptophyta as well as bacterial and fungal PALs. In agreement with previous studies (Emiliani *et al.*, 2009; de Vries *et al.*, 2017), the fungal and bacterial PAL sequences are closely related (bacterial clade: bootstrap support 100; fungal clade: bootstrap support 85) to the clade of land plant PALs (bootstrap support: 96). Additionally, we included diverse eukaryotic and prokaryotic HALs based on the set obtained from de Vries *et al.* (2017) which, as in this latter study, form two clades with eukaryotic and prokaryotic HALs (bootstrap support 83 and 97). All putative PAL-like candidate sequences from *C. braunii* clustered with HAL sequences, one (*C. braunii* g66119_t1) with a cyanobacterium (bootstrap support 100), and two (*C. braunii* g34530_t1 and *C. braunii* g57646_t1) sister to an entire bacterial HAL clade (bootstrap support 100). Querying *C. braunii* g34530_t1 in I-TASSER recovered, again, the *P. putida* HAL with a very good TM-score (0.914; 1GKM; Baedeker and Schulz, 2002). Residues involved in substrate binding map to a hypothetical binding pocket predicted by I-TASSER—including the prosthetic group MIO; as with the aforementioned *Klebsormidium* PAL homolog, residues present in the C-terminal region of *C. braunii* g34530_t1 could contribute to the active site if g34530_t1 forms a homotetramer (Figures S1c and 2). In contrast to the phylogenetic position of the sequences from *C. braunii*, both sequences retrieved for *K. nitens* clustered with the PAL clades, one *K. nitens* kfl00024_0250_v1.1 with low support (bootstrap 61) and a rather long branch as sister to plant and fungal PALs, showing that its placement is not fully resolved and that further analyses are required to identify its true identity. The second *K. nitens* sequence, kfl00104_0290_v1.1, clustered within the clade of bacterial PALs (bootstrap support 100), of which some were already functionally characterized—for example, the characterized PAL of *Nostoc punctiforme* (Moffit *et al.*, 2007). This is in agreement with the placement of this protein sequence in de Vries *et al.* (2017), and supports it as a putative PAL sequence.

In sum, the evolutionary origin of streptophyte PAL appears to be complex and remains obscure: it may be that PAL had a distinct origin in streptophyte algae and land plants, yet the pattern may also be explained by an origin via an endosymbiotic gene transfer from the cyanobacterial plastid progenitor that was retained by streptophytes with a gain of an extra C-terminal domain later in the evolution of embryophytes, resulting in the two

distinct PAL clades. It is, however, important to note that the 3'-region of the genomic locus that codes for the shorter *K. nitens* protein kfl00104_0290_v1.1 contains sequence information that resembles code for the missing C-terminal stretch; thus, the C-terminal stretch might have simply been secondarily lost in *K. nitens*. Independently, the presence of PALs in fungi further complicates the evolutionary scenario. Rampant gene losses during eukaryotic evolution or convergent domain acquisitions, as well as horizontal gene transfer (as hypothesized by Emiliani *et al.*, 2009) are other scenarios that can explain the evolutionary origin of PALs in streptophytes and thus ultimately in land plants.

Streptophyte algae have an expanded and divergent repertoire of cytochrome P450 monooxygenases with no clear C4H orthologs

After the synthesis of cinnamate by PAL, two routes open up (Figure 1). One of them is the conversion of cinnamate into *p*-coumarate, which is catalyzed by cinnamate 4-hydroxylase (C4H). C4H belongs to the large class of CYP450 enzymes, which are present in all domains of life (Omura, 1999). Among the CYP450 enzymes, C4H belongs to the CYP450 subfamily 73 (CYP73). In land plants, CYP450s have undergone substantial duplication and subfunctionalization, underpinning the specialized metabolic capabilities of embryophytes (Nelson and Werck-Reichhart, 2011); for example, the CYP73 subfamily belongs to the larger CYP71 clan. The specific CYP450 monooxygenases that fall into the group of C4H appear to be limited to land plants: clear orthologs can be found in bryophytes and tracheophytes (Emiliani *et al.*, 2009; de Vries *et al.*, 2017). That said, the product of the reaction carried out by C4H in land plants (*p*-coumarate) has been detected via UHPLC-MS/MS in phylodiverse algae (Goiris *et al.*, 2014). Therefore, there appears to exist a route towards *p*-coumarate that is either independent of C4H via direct transformation of tyrosine by PTAL or, for example, carried out by a highly divergent C4H homolog. PTALs have so far, however, been observed in monocots (Barros and Dixon, 2020; Barros *et al.*, 2016), suggesting a different CYP73 subfamily enzyme that may carry out the reaction. Owing to the recent increase in genomic data available for streptophyte algae, we revisited the question of when C4H-based *p*-coumarate might have emerged and explored CYP450 evolution.

We sampled C4H homologs from seven land-plant genomes that had a BLAST bit score (a normalized alignment score) of at least 200, as well as seven streptophyte algal and seven chlorophyte algal genomes that had a bit score of at least 100. We aligned all C4H homologs and computed a maximum likelihood phylogeny (Figure S2). The well-characterized C4H of *Arabidopsis* fell into a clade with full (100) bootstrap support; this clade included at least

one C4H homolog from each of the other six land-plant genomes, corroborating the notion that all land plants have C4H orthologs, which appear conserved in their function in both bryophytes and tracheophytes (Ro *et al.*, 2001; Russel and Conn, 1967; Urban *et al.*, 1994; Wohl and Petersen, 2020) and thus since the LCA of land plants. However, no algal sequences fell into this clade. That said, we observed four well-supported clades of streptophyte algal CYP450 enzymes (Figure S2). Investigating the genetic distances, we find that some of the streptophyte algal sequences have a closer genetic distance to the C4H-like clade than to sequences from land plants (including *Arabidopsis thaliana*) from other CYP450 subfamilies (Table S1). While these sequences remain of unknown function, they are candidates for the CYP450 enzyme family that catalyzes the C4H-function in algae.

A deep split of streptophyte 4CL/ACS

The second route that opens up after the PAL-dependent step is the conversion of cinnamate into cinnamoyl-CoA. This is carried out by the AMP-forming synthetase/ligase 4CL and potentially other enzymes annotated as acyl-CoA synthetases (ACS/ACoS; Shockey *et al.*, 2003; Figure 1). Altogether, these enzymes belong to a large family of distantly related acyl-activating enzymes (AAEs), such as the long-chain acyl-CoA synthetases (LACS) and many more (Shockey *et al.*, 2003; Figure S3). Homologs with affinity to 4CL appear to occur across chlorophytes and streptophytes (Labeeuw *et al.*, 2015). At least in *Arabidopsis*, the family of 4CLs has expanded and includes four canonical ('4CL') and nine additional 4CL-like ('4CLL') members, falling into AAE clade IV and V as defined by Shockey *et al.* (2003). We thus set out to understand what the 4CL repertoire of the LCA of land plants and the one shared with algae might have looked like.

In order to trace the radiation of 4CLs across the green tree of life, we sampled 4CL homologs from genomes of nine land plants, seven streptophyte algae (plus four transcriptomes of streptophyte algae) and seven chlorophyte algae that had a minimum of 400 and a maximum of 1150 amino acids in length and showed affinity to the 4CL clade in a larger phylogenetic survey (Figure S3). We recovered a large clade (bootstrap support 85) that included all *bona fide* 4CL paralogs and ACOS5 of *A. thaliana* (Figure 3); ACOS5 has been previously associated with the *bona fide* 4CL clade (Shockey *et al.*, 2003), but it did show only inconsistent activity on typical substrates of 4CL (Costa *et al.*, 2005) and appears to have a very specific function in sporopollenin biosynthesis of pollen (de Azevedo Souza *et al.*, 2009). These observations agree with the presence of different (but conserved) amino acids at sites that bind hydroxycinnamate in typical 4CLs (Figure 3), which might suggest a different natural substrate for ACOS given that affinity is mostly determined by the binding pocket size (Hu *et al.*, 2010). We

further recovered the angiosperm-specific separation defined by Ehltting *et al.* (1999) into class I and class II 4CLs. Additional lineage-specific radiations occurred, for example in *P. patens*, which fell into a clade of bryophyte sequences (bootstrap support of 70), and in *Selaginella moellendorffii* (spread out over the fully-supported clade of 4CLs). The common ancestor of land plants appears to have possessed an ACOS5-like and one 4CL-like gene, all other 4CL paralogs in this clade likely emerged later during land plant evolution. Clustering with AAE clade IV [including *AtACOS5*, and *At4CL1,2,3* and 5 (bootstrap support 85)] are sequences from five streptophyte algae. Each of the five streptophyte algae possesses one homolog to these five types of AMP-forming ligases with 4-coumarate-CoA synthesizing activity. When we predicted the tertiary structure of *C. atrophyticus* Chrsp175S02417 and *Penium margaritaceum* 006213.t1 via I-TASSER (Zhang, 2008) we recovered, in both cases, firefly luciferases as best match (TM-scores 0.852 and 0.918; 1BA3 and 2D1S; Franks *et al.*, 1998; Nakatsu *et al.*, 2006). Investigating the putative structure of the other streptophyte algal sequences (*Zygnema circumcarinatum* DN42558_c0_g1_i1, *Spirogyra pratensis* 3442_c2_g1_i6 and *K. nitens* 00016_0470_v1.1), however, always recovered *Populus tomentosa* 4CL (3A9U; Hu *et al.*, 2010) as their closest structural analog (TM scores of 0.957, 0.969, and 0.961, respectively). Hence, we hypothesize that a 4CL/ACOS5-like encoding gene was present in the LCA of all streptophytes. Underpinning this hypothesis is that the sequence of the amino acids in the binding pocket in the streptophyte algal 4CL homologs is consistent with that of 4CL homologs from other land plants (including that of *A. thaliana*; Figure 3). Further, the amino acids relevant for the enzymatic function (i.e. the residues KQK involved in adenylation, nucleophilic substitution and coumaroyl-AMP cleavage) are also conserved across most 4CL/ACOS5 sequences, including those of the streptophyte algae. Variation in these residues is already apparent in 4CLL homologs, and outside of the 4CL/ACOS5/4CLL clade these residues show high variability (Figure 3).

A similar pattern was observed when we investigated the domain structure of all recovered sequences. Most of the 4CL/ACOS5-like sequences contained four domains: phosphopantetheine binding ACP domain (IPR025110), AMP-binding, conserved site (IPR020845), AMP-dependent synthetase-like superfamily (IPR042099), AMP-dependent synthetase/ligase (IPR000873; Figure S4). There were four exceptions to this pattern. They include one sequence from the water fern *Azolla filiculoides* (Azfis0013.g013344) and two hornwort sequences from *Anthoceros agrestis* BONN (Sc2ySwM344.2803.3 and Sc2ySwM344.2803.4), which all missed the phosphopantetheine binding ACP domain (IPR025110). The other exception was *AtACOS5*, which is the only sequence in this clade that missed the conserved

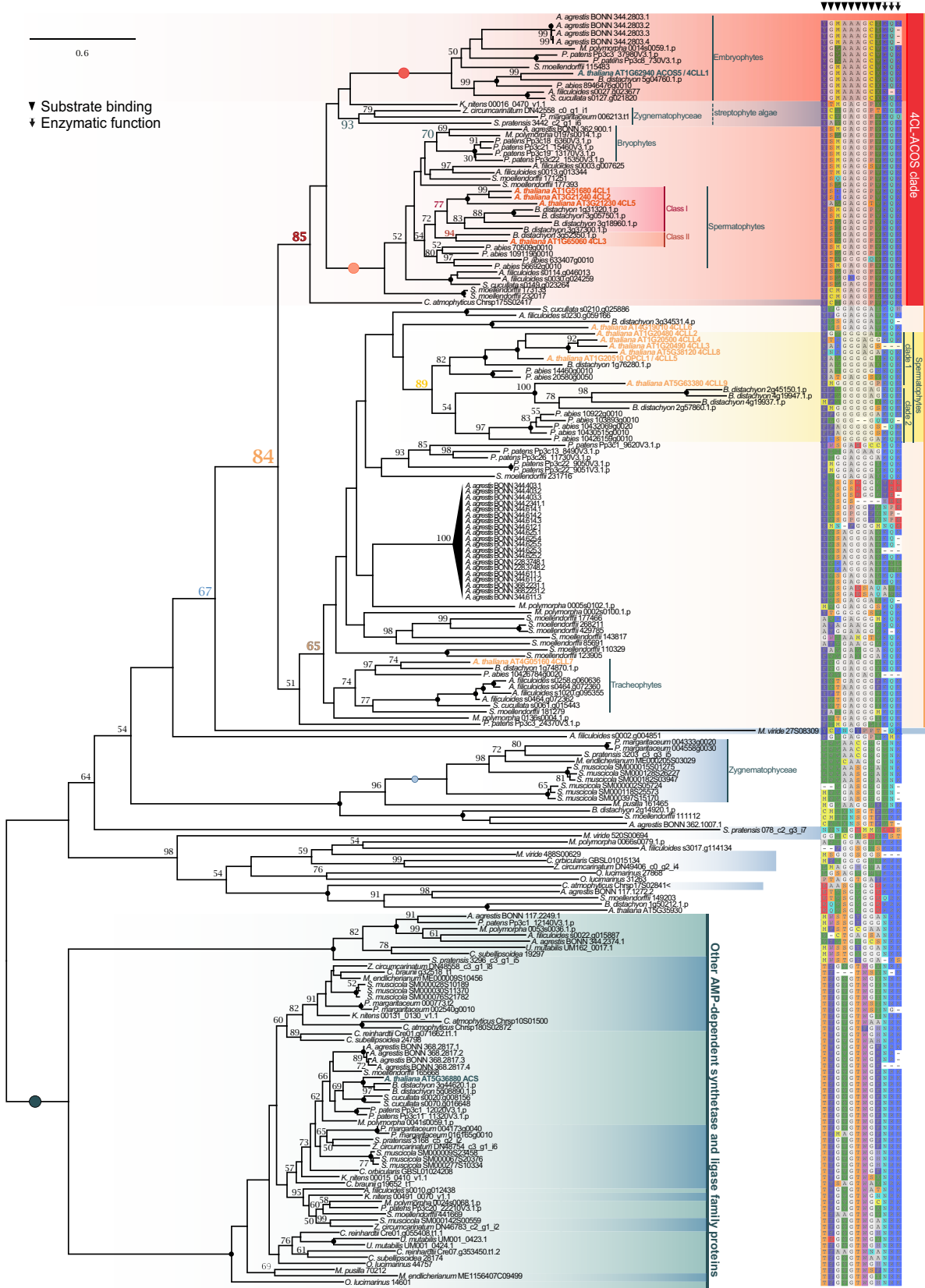


Figure 3. 4-Coumarate-CoA ligase (4CL) homologs occur across Streptophyta.

4CL homologs were sampled from protein data of nine land plants, seven streptophyte algal and five chlorophyte algal genomes. Only protein sequences with a minimum length of 400 and a maximum length of 1150 amino acids were included. From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny. On the right we show key residues for substrate binding and function of canonical 4CL as reported by Hu *et al.* (2010); the clade(s) of bona fide 4CLs were assigned the warmest/reddish colors.

AMP-binding site (IPR020845). The domain pattern is similar across the 4CL-like clade, too. Yet, more sequences miss either the IPR025110 and/or the IPR020845 domain. The streptophyte algal sequences within the ACOS5/4CL clade contained all four domains, whereas algal sequences outside of this clade missed at least one—but recovered several other domains. These additional domains are not conserved within the phylogenetic subclades of these algal sequences and only exceptionally occur in the 4CL/ACOS4/4CLL (two sequences) or other AMP-dependent synthetase and ligase family protein clades (two sequences).

We recovered a second clade of spermatophyte sequences (bootstrap support 89) representing AAE clade V enzymes, which contains 4CLL8 and several other ATP-ligases of *A. thaliana* with predicted 4CL activity, including OPCL1 (matching 4CL-like 5 with a 100% amino acid identity according to Uniprot). Sequences in this clade, however, diverge in the amino acids that are involved in the formation of the binding pocket in the canonical 4CLs (Figure 3), which might point to a different substrate preference of the enzymes in this clade. In fact, OPCL1 and many of these '4CLs' (e.g. AT5G63380, AT1G20500, AT4G05160) showed higher activity on fatty acids and fatty acid-derived precursors for the phytohormone jasmonic acid than cinnamate-derived compounds in an *in vitro* substrate survey carried out by Kienow *et al.* (2008). It is thus questionable that the enzymes of this clade act as *bona fide* 4CLs. Homologs to these sequences are found in *Brachypodium distachyon* and *Picea abies*, suggesting an origin in the LCA of seed plants followed by two duplication events, with either: (a) both taking place in the LCA of angiosperms; or (b) one early on in the LCA of seed plants and the second in the LCA of angiosperms. Bootstrap support to include the *P. abies* sequences in the clade containing AT5G63380 (4CLL9) is, however, low (bootstrap 54). Each duplication event was followed by independent lineage-specific radiations giving rise to a whole plethora of possible candidates for 4CL, but also a large evolutionary potential with regard to substrate specificity and flexibility. The 4CLL clade of spermatophytes is nested within a larger, lowly supported clade (bootstrap 65) that included sequences from across the land-plant tree of life. Here, pronounced and independent expansion occurred in most of the major lineages of land plants, leading to large clades of, for example, proteins of the hornwort *Anthoceros* and the lycophyte *Selaginella*. As noted above, most of the *Selaginella* and some of the *Anthoceros* homologs

retained the conserved KQK residues required for the catalytic activity but others did not (Figure 3), which suggests the presence of species-specific functions. Outside of the entire 4CL-ACOS and 4CLL clade ('Streptophyte 4CL/ACOS/4CLL-likes'; bootstrap 84) clustered various highly divergent ATP-dependent synthetases and ligases that exist throughout the green tree of life including sequences from chlorophytes. None of these synthetases and ligases retained the catalytic triad KQK.

Altogether, our phylogenetic data indicated that a 4CL/ACOS5-like encoding gene was present at the base of Streptophyta. Domain annotation and the analysis of amino acid patterns in the binding pocket and functional sites support this idea. Further, the similarity of these residues between the candidates of streptophyte algal homologs for 4CLs and the sequences of 4CL proteins with high activity on cinnamate derivatives as substrates (Costa *et al.*, 2005) indicates that 4CL activity may have evolved more than 700 million years ago in streptophytes.

Patchy distribution of CCR-like sequences in streptophyte algae and pronounced independent radiations in land plants

En route to the production of different lignin monomers is the NADPH-dependent reduction of the activated acyl-group of the phenylpropanoid backbone molecules. This first step towards an aldehyde functionality is carried out by CCR (Figure 1), which falls into a larger family of NADPH-dependent reductases, including dihydroflavonol reductases (DFRs) and DFR-likes (DFL; Devic *et al.*, 1999; Lacombe *et al.*, 1997). We previously reported the presence of CCR-like protein sequences in streptophyte algae (de Vries *et al.*, 2017). Since these previous analyses, however, genome data on additional major lineages of land plants and streptophyte algae have become available.

With these more exhaustive data at hand, covering most major lineages of streptophyte algae and all major lineages of land plants, it is now possible to infer the evolutionary history of CCR-like and DFR-like sequences. We computed a maximum likelihood phylogeny of CCR homologs with a minimum of 220 amino acids that we detected in genomes of 15 land plants, seven streptophyte algae and seven chlorophytes; additionally, we included sequences found in the transcriptomes of the Zygnematophyceae *S. pratensis* (de Vries *et al.*, 2020), *Z. circumcarinatum* (de Vries *et al.*, 2018), and the Coleochaetophyceae *Coleochaete orbicularis* (Ju *et al.*, 2015; Figure 4).

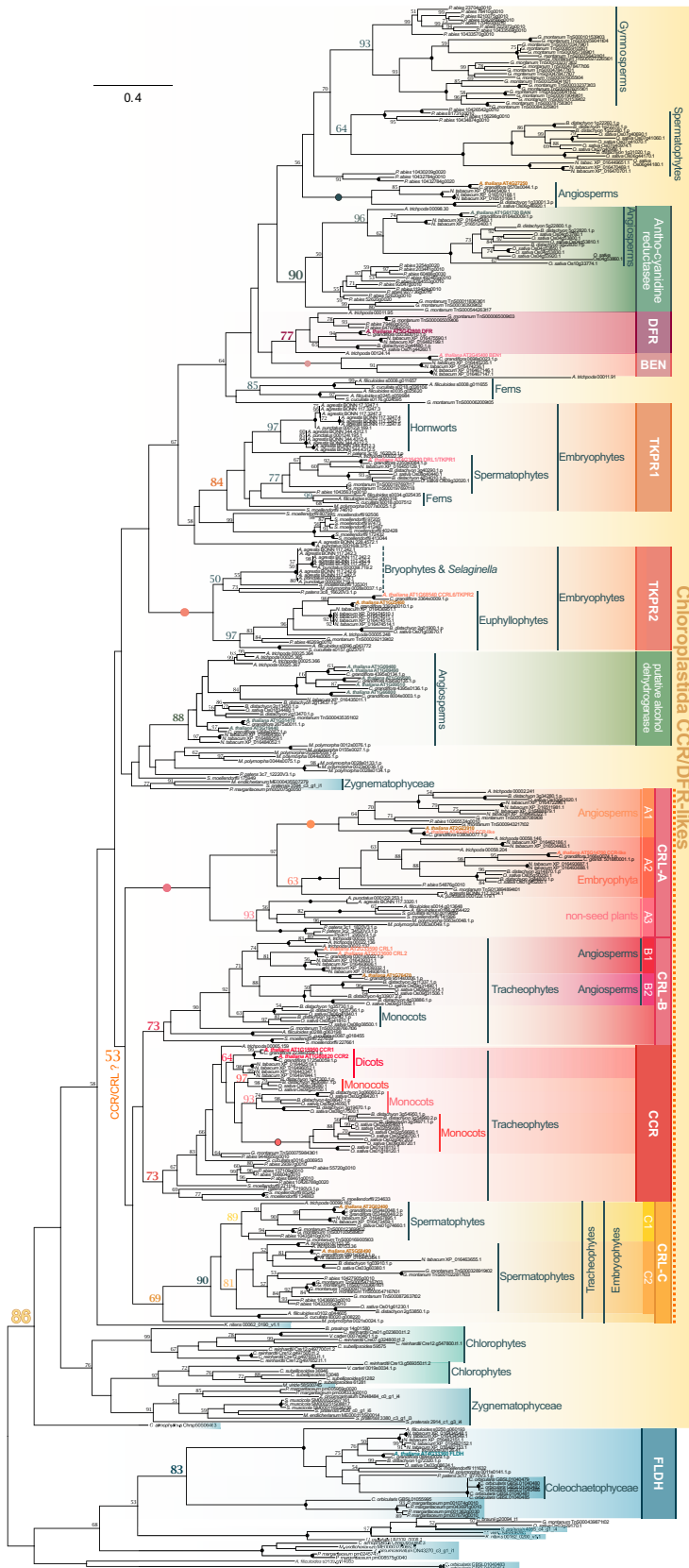


Figure 4. The complex evolutionary history of cinnamoyl-CoA reductase (CCR) in Chloroplastida. CCR homologs of a minimum of 220 amino acids were sampled from protein data of 15 land plants, seven streptophyte algal and eight chlorophyte algal genomes; additionally, we included sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *Coleochaete orbicularis* (Ju *et al.*, 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; the clade(s) of *bona fide* CCRs were assigned the warmest and brightest reddish colors.

1365313x, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.15387 by Georg-August-Universität Göttingen, Wiley Online Library on [25/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The CCR homologs were distributed over several major clades. This included the CCRL/DRL-like sequences described as TETRAKETIDE α -PYRONE REDUCTASE (TKPR) by Grienenberger *et al.* (2010), which is an important enzyme acting in the production of sporopollenin. We recovered a well-supported clade of TKPR1 homologs (bootstrap support of 84) and fully supported clade of TKPR2 homologs. Both clades of TKPRs contained homologs from across the diversity of land plants, bolstering the idea that TKPR1 and TKPR2 split early during plant evolution (Grienenberger *et al.*, 2010)—before the most recent common ancestor of land plants came about. Our domain structure analyses found that TKPR1 possessed the NAD-dependent epimerase/dehydratase (IPR001509) and NAD (P)-binding domain superfamily (IPR036291) domains, which appear to be present in most sequences included in the phylogeny as well as the Tetraketide alpha-pyrone reductase 1 (IPR033267) domain (Figure S5). In contrast, TKPR2 only encoded the first two domains, which is more similar to what is found in the CCR clade.

The *bona fide* CCRs and CCR-like were spread out over two clades. These two clades were nested in a weakly-supported monophylum (bootstrap support of 53), which was sub-divided into four medium to fully supported clades. A fully supported clade of CCR-like (including AT4G30470 and AT2G23910) included sequences from across embryophytes; we coined this monophylum CRL-A. *AtCCRL1* and *AtCCRL2* appear to be co-orthologs to one sequence in the Brassicaceae *Capsella grandiflora* (0380s0077.1.p), thus our data suggest a limited distribution of direct orthologs to CCRL1 and CCRL2. Yet the two sequences together fall into a large clade, here coined CRL-B (bootstrap support 73), that contained sequences from all major lineages of tracheophytes. Another medium-supported clade (bootstrap support 73) included the *bona fide* CCRs, CCR1, and CCR2 of *Arabidopsis*. The duplication that resulted in these two CCRs occurred earliest in the common ancestor of all rosids and latest in the common ancestor of Brassicaceae, yet the CCR clade included homologs from across tracheophytes. Many of these lineages appear to have expanded their own repertoire from one CCR1/2 homolog that was present in the LCA of tracheophytes. Interestingly, a clade of divergent monocot CCRs display several replacements in key amino acids involved in the binding of the substrate's phenolic ring, in particular a replacement of non-polar aliphatic Ile to aromatic Tyr/Phe (Figure S6), which might reflect a difference in substrate affinity. The catalytic triad SYK (Figure S6; Pan *et al.*, 2014) is required for enzymatic activity, and is overall conserved across CCR/DFR-like and FLDHs. Interestingly, several CRL-As possess non-conservative amino acid replacements from large phenolic (Tyr/Phe) to smaller (His, Leu, Ser, Gly) amino acids, which might suggest divergent substrate affinities for CRL-As. This is

consistent with the domain structure of many CRL-A sequences, which often lack the NAD-dependent epimerase/dehydratase (IPR001509) domain, but possess additional domains such as 3- β -hydroxysteroid dehydrogenase/isomerase (IPR002225) or match an additional NAD(P)-binding domain (IPR016040; Figure S5). This pattern is only occasionally occurring in sequences from the CCR or other CRL clades.

Finally, there is a third clade with a bootstrap support of 69 that included tracheophyte sequences (forming a sub-clade with a bootstrap value of 90) and a single *M. polymorpha* homolog; we coined this clade CRL-C. Altogether, this suggests that the LCA of all land plants had two homologs of CCRs/CRLs: one CRL-A and one CRL-B, CRL-C or CCR homolog. In vascular plants, duplications have resulted in sub-clades of the CRL-B/CRL-C/CCR homologs.

Within the larger clade that encases the DFRs, DFRLs, CCRs and CCRLs ('Chloroplastida CCR/DFR-like'; bootstrap 86), one supported clade of Zygnematophyceae (bootstrap support 77) and one supported clade of chlorophyte and streptophyte algae (bootstrap support 76) exists. This points to a distinct DFR/DFRL/CCR/CCRL clade that arose in the ancestor of Zygnematophyceae, yet its placement within the phylogeny other than it belonging to the larger DFR/DFRL/CCR/CCRL clade is uncertain. The divergent pattern of amino acids, which perform substrate and cofactor binding in land plants, suggest that these algal homologs might vary in the substrate and enzymatic activity compared with plant CCR/DFRs (Figure S6). It is, however, certain that within land plants, a pronounced radiation of CCRs occurred.

CAD homologs are present across the green lineage

The second reduction step of the activated acyl-group of the phenylpropanoid backbone and one of the last steps in lignin biosynthesis is the production of phenylpropanoid-derived alcohols from the corresponding aldehydes. An example is the conversion of *p*-coumaroyl aldehyde into *p*-coumaryl alcohol (Kim *et al.*, 2004; Pan *et al.*, 2014). The required reduction is catalyzed by CAD (Figure 1), which is the rate-determining enzyme by which lignin is produced (Gross *et al.*, 1973; Mansell *et al.*, 1974). In *A. thaliana*, there are at least 11 enzymes belonging to the CAD family. Enzymes of the CAD family have been divided into five major groups, of which group IV was described as monocot-specific (Saballos *et al.*, 2009). In our previous studies, many of the chlorophyte and streptophyte potential CAD homologs, identified mostly from transcriptomes and few genomes of algae, were described as CAD-like or CAD group II/III-affiliated (de Vries *et al.*, 2017, 2020). Sequences clustering with those of CAD group II have been characterized as sinapyl alcohol dehydrogenase (SAD) or show predicted structural similarity to SAD enzymes (Guo *et al.*, 2010; de Vries *et al.*, 2017).

Additionally, some SADs appear involved not in the synthesis of lignin but defense compounds such as lignans (Barakate *et al.*, 2011; Guo *et al.*, 2010; Saleem *et al.*, 2010; Suzuki and Umezawa, 2007), and it may thus be that CAD group II is functionally versatile.

Here, we used the 11 canonical CAD sequences to understand the diversity in CAD homologs across streptophytes. This includes also CAD group II sequences, for which homologs in other species may have other substrate specificities and thus are involved in different steps of the phenylpropanoid pathway (Barakat *et al.*, 2009; Guo *et al.*, 2010). We computed a phylogeny of CAD homologs (Figure 5) detected in phylo diverse Chloroplastida. While the resolution of the backbone is weak, we recovered all five CAD groups defined by Saballos *et al.* (2009). All the CAD groups were resolved as land plant-specific clades of CAD homologs with robust support. Each clade contained a varying set of major land plant lineages (described below); the clades of putative streptophyte algal CAD homologs contained both fewer proteins and fall in-between the five CAD groups. Hence, this more phylo diverse dataset tells a more complicated evolutionary history for CAD homologs than the less-diverse data from de Vries *et al.* (2017).

Our data suggest that the common ancestor of Zygnematophyceae and land plants may have possessed two CAD-like genes, which was followed by lineage-specific radiations. While it appears—based on the overall topology of the tree—appealing to suggest that one gene gave rise to CAD-group V and the other ancestral gene was the basis of CAD-groups I–IV, the low statistical support for the backbone of the phylogeny does not allow to confirm such a hypothesis (Figure 5). We can infer that the earliest land plants likely inherited a few (or just one) CAD homolog from their algal progenitors. Most of the radiation of CADs occurred in plants dwelling on land. Of the canonical CAD group—and based on those parts of the topology with high bootstrap support—CAD-group V (containing AtCAD1) is the only group present in all major land plant lineages (Figure 5). CAD-group I was likely present in the LCA of tracheophytes, as it includes sequences from angiosperms, gymnosperms, ferns and the lycophyte *S. moellendorffii*. CAD-groups II, III and IV include only angiosperm sequences—but note that with very weak bootstrap support (53) a sequence from the gymnosperm *Gnetum montanum* associates with the group CADII/III/IV; likely, expansion resulted in the ancestral gene of CAD-groups II and III, which diverged into CAD group III genes in angiosperms and after another expansion CAD-group II originated in the LCA of dicots. This is in contrast to de Vries *et al.* (2017), where the streptophyte algal CAD-like sequences were clustering with CAD group II/III sequences, but resembles the placement of transcriptomic CAD-like sequences from *S. pratensis* and *Mougotia* sp. in an already more diverse phylogenetic analysis (de Vries *et al.*,

2020). This is a clear case where including a larger diversity of streptophyte sequences to the analysis enables us to better understand the complexity of the evolution of highly radiated gene families. An analysis of the residues salient to CAD function showed a general conservation of residues involved in zinc and NADP⁺ binding across Chloroplastida CAD-like sequences (Youn *et al.*, 2006), whereas the residues in the binding pocket are generally less conserved in these sequences (Figure 5). However, within a canonical CAD-group or a CAD-like clade we see a general conservation of residues in the binding pocket. The binding pockets of each CAD-like clade appear different than those of the canonical CAD-groups. Domain analyses suggest that some of the CAD-like sequences of streptophyte algae do not encode all of the five domains present in most canonical CAD sequences, yet many of the Zygnematophyceae CAD-like sequences encode all of these five domains. Similar patterns emerge for other CAD-like sequences (Figure S7).

All CAD groups are shaped by multiple lineage-specific duplications and losses. This hampers the inference of function and substrate specificity of the diverse CAD-like sequences. Additionally, several lineages have originated a variety of CAD homologs that are not yet designated to previous groups. In the absence of functional data, we, however, will not give them a group designation but rather designate them as lineage-specific CAD-homologs of unknown SAD or CAD function. CAD-like homologs found in streptophyte algae show similar functional residues to other CAD-like of land plants (not included in any of the five clades of canonical CADs), including at the binding pockets—the pattern of residues is similar to what is observed for other land plant sequences in-between the *bona fide* CAD groups. The *bona fide* CAD groups showed a more homogeneous pattern of functional residues.

Overall, both the topology of the phylogenetic tree and the conservation of key residues point to: (i) a deep evolutionary origin of CAD homologs; and (ii) independent radiations of CADs—not only in land plants but also in streptophyte algae.

Independent radiations of acyltransferases and scattered candidates in streptophyte algae

A versatile group of enzymes that are important for the processes leading up to the lignins but also compounds with antioxidant and antimicrobial properties are the BAHD acyltransferases (named after the first enzymes characterized for this family BEAT, AHCT, HCBT and DAT; D'Auria, 2006). Most prominent among these are the versatile hydroxycinnamoyl-CoA shikimate/quinic hydroxycinnamoyltransferases (HCT; Eudes *et al.*, 2016). Recently, Kriegshauser *et al.* (2021) reported on the functional conservation of the HCT homologs found in bryophytes with those of seed plants.

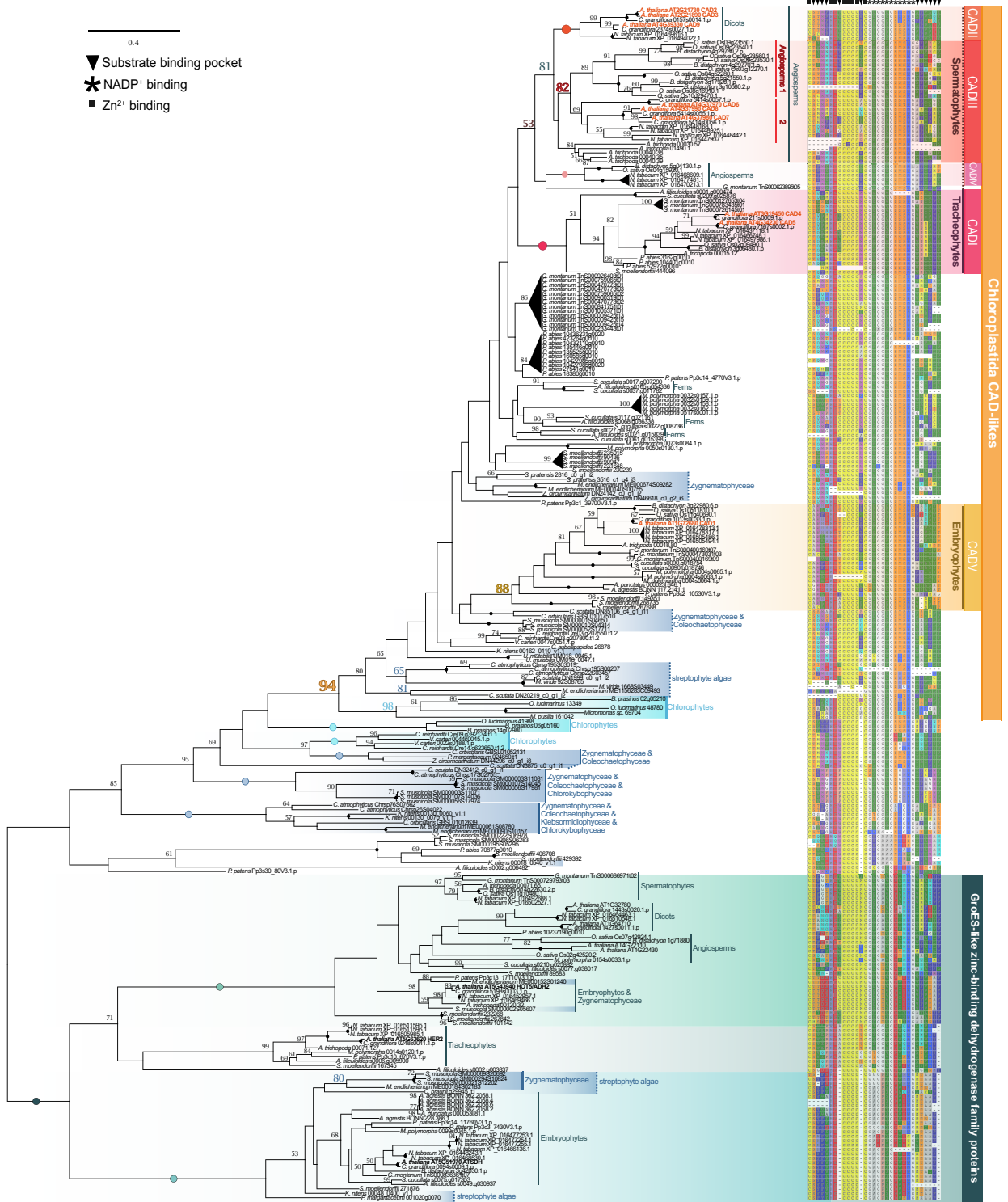


Figure 5. Phylogenetic analysis highlights cinnamyl alcohol dehydrogenase (CAD) candidates across Chloroplastida. CAD homologs were sampled from protein data from 15 land plants, seven streptophyte algal and eight chlorophyte algal genomes, as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Coleochaeta scutata*, *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *C. orbicularis* (Ju *et al.*, 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; the clade(s) of *bona fide* CADs were assigned the warmest/reddish colors. The five groups of CADs were named in accordance with Saballos *et al.* (2009). Next to the sequence labels residues from the binding pocket, NADP⁺- and Zn²⁺ binding are shown—based on Youn *et al.* (2016).

Figure 6. A clade of 4-coumarate 3-hydroxylase (C3H) orthologs originated at the base of land plants.

C3H homologs were sampled from protein data of genomes of 15 land plants, seven streptophyte algae and five chlorophytes; additionally, sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *Coleochaete orbicularis* (Ju *et al.*, 2015) were included. For downstream analyses, we used either: (a) all sequences that had a bit score of at least 100; or (b) the top five hits. We aligned all sequences, cropped them to the alignable region, and an unrooted maximum likelihood phylogeny was computed using LG+F + I+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; the clade(s) of *bona fide* CYP98As were assigned the warmest and brightest reddish colors. Two large clades that contained only: (i) *Anthoceros*; and (ii) chlorophyte and streptophyte algal sequences were collapsed; the full tree is shown in Figure S13.

We computed a phylogeny including phylo diverse acyltransferases (Figure S8). Overall, the topology of the tree corroborates the findings of Kriegshauser *et al.* (2021) that a clear clade of HCT proteins first emerged in land plants—streptophyte algal sequences were few, divergent from HCT, and scattered over the tree without clear affinity to characterized acyltransferases. Without functional analyses, there is no solid foundation for predicting their function—making them exciting candidates for future studies. What our data, however, clearly reveal is that the LCA of land plants likely had an expanded repertoire of acyltransferases that further diversified during the radiation of plants on land.

A clear clade of C3Hs is limited to land plants

At several steps of the phenylpropanoid pathway, enzymes belonging to the cytochrome P450 family CYP98A, within the large CYP71 clan (Nelson and Werck-Reichhart, 2011), carry out hydroxylations of *p*-coumarate-derived compounds (such as *p*-coumaroyl esters; for function of C3H see Figure 1). In land plants, these hydroxylations are important for the production of lignins, lignans, volatile phenylpropanoids, coumarins and many more phenylpropanoid-derived compounds. Previously, de Vries *et al.* (2017) described the detection of C3H in all land plants and one putative C3H ortholog in *K. nitens*. Now, with genomic gaps in the streptophyte tree of life filled, we revisited the distribution of C3H.

The number of C3H homologs detected in genomes of 15 land plants, seven streptophyte algae and seven chlorophytes, and in the transcriptomes of *S. pratensis* (de Vries *et al.*, 2020), *Z. circumcarinatum* (de Vries *et al.*, 2018) and *C. orbicularis* (Ju *et al.*, 2015) varied strongly between lineages. When sampling the sequences via BLAST (with AT2G40890 as query sequence), we thus included either: (a) all sequences that had a bit score of at least 100; or (b) the top five hits. We aligned all sequences, cropped them to the alignable region and computed a maximum likelihood phylogeny (Figure 6).

A clade of CYP98A included sequences from all major lineages of land plants. This suggests that at least one CYP98A sequence was present in the LCA of all land plants. Based on the lineages included here, it appears that from an ancestral single copy gene, radiations occurred in the dicot lineages and *Amborella trichocarpa*. A single

copy remained in the bryophytes, lycophytes, ferns, gymnosperms and monocots. The clade of CYP98A8 and CYP98A9 was in our dataset limited to the Brassicaceae *A. thaliana* and *C. grandiflora* (bootstrap 100). These two enzymes function in a route derived from the phenylpropanoid pathway and are involved in the formation of N^1, N^5 -di(hydroxyferuloyl)- N^{10} -sinapoylspermidine (Matsuno *et al.*, 2009). The CYP98A8/9 clade falls into the larger CYP98A clade (bootstrap support 100) together with the C3H sequences, suggesting that they are the closest paralogs of C3H in *A. thaliana* and *C. grandiflora*. Analyses of their substrate recognition sites (SRSs; Rupasinghe *et al.*, 2003) support the divergent functional roles between CYP98A8/9 and canonical C3H. In particular, the first two SRSs (SRS1 and SRS2) show various distinct amino acid differences between the CYP98A8/9 clade and the C3H clade (Figure S9). A similar pattern is observed for the clade CYP98A enzymes in non-seed plants (Figure 6; bootstrap support 79). This clade includes the sequences from *P. patens* (Pp3c22_19010V3.1.p) and *S. moellendorffii* (271465) that have been shown to have their highest activity on 4-coumaroyl-anthranilate instead of 4-coumaroyl-shikimate and -quinic acid (as it is the case in *A. thaliana* and other angio- and gymnosperms; Alber *et al.*, 2019). Here, too, SRS1 and 2 show several differences as well as more variability than the C3H clade of angiosperms (Figure S9). SRS1 and 2 are predicted to be involved in binding of the substrate tails, which exhibit strong variation in size (Rupasinghe *et al.*, 2003), and thus may be critical for the substrate specificity of these paralogs. Indeed, SRS1 and 2 are the two SRSs showing the strongest variation across the entire phylogeny, including also other CYP450 subfamilies (Figure 6), corroborating this hypothesis.

We recovered additional land-plant-specific clades of CYP450 enzymes (Figure 6), such as one containing TRANSPARENT TESTA7 (TT7)-like sequences (bootstrap support 92); TT7 is a cytochrome P450 75B enzyme required for flavonoid 3' hydroxylase activity in the flavonoid biosynthesis (Schoenbohm *et al.*, 2000; Tanaka *et al.*, 1997). The BLAST search used for sampling C3H homologs further recovered *AtCYP71B* and *AtCYP76C* members. In our phylogenetic analysis, we inferred that *AtCYP71B34* and *AtCYP71B35* were likely born out of an *Arabidopsis*-specific duplication, while CYP71B enzymes in general are present across angiosperms. In contrast, *AtCYP76C1* and

AtCYP76C4 appear to have originated prior to the split of *A. thaliana* and *C. grandiflora*; a CYP76C4 ortholog appears to have been lost in the latter plant species. The CYP76C clade may also be represented in other species outside of angiosperms, because we found a sequence from *G. montanum* clustering with these sequences with a bootstrap support of 87. The rather long branch warrants attention and would require a more CYP76C-focused phylogenetic analysis, which is not the point of this paper.

The algal sequences showed strong divergence to the C3Hs, forming only a larger (fully supported) streptophyte-specific clade with all the recovered and functionally diverse CYP450 enzymes (Figure 6). Domain structures of these streptophyte algal sequences are the same as for the C3H homologs, but are in general conserved across the phylogeny independent of the CYP450 subfamily assignment (Figure S10). Only a few scattered exceptions occur. Given the low support of most of the tree backbone, the role(s) of the streptophyte algal homologs detected here remains elusive. Most of the streptophyte algae show independent radiations of their CYP450 enzymes complicating functional predictions even further. Thus, while there are interesting CYP450 candidates in streptophyte algae, a clear C3H clade likely first arose early during the evolution of embryophytes.

MAGLs: multiple early radiations, independent subfunctionalization, and the origin of caffeoyl-5-*O*-shikimate esterase (CSE)

The conversion of caffeoyl-5-*O*-shikimate to caffeic acid may be a step along the biosynthetic routes that lead to the production of G- and S-lignins in certain vascular plants (Figure 1). The enzyme responsible for this step is CSE. CSE converts caffeoyl-5-*O*-shikimate to caffeic acid, and was hypothesized to act together with 4CL/ACOS5 to circumvent the catalysis of caffeoyl-5-*O*-shikimate to caffeoyl-CoA via HCT (Vanholme *et al.*, 2013). The latter pathway was proposed for tobacco by Hoffmann *et al.* (2003), and confirmed for *A. thaliana in vitro* by Vanholme and colleagues (2013). Yet, based on *cse* mutants in *A. thaliana*, they suggested that synthesis of caffeoyl-CoA is more likely to occur via CSE and 4CL/ACOS5 than directly from caffeoyl shikimate by HCT *in planta*. That said, in the model grasses *B. distachyon* and *Zea mays*, no CSE orthologs are present and crude extracts from these species show little signs for the characteristic esterase activity (Ha *et al.*, 2016). On the other hand, non-vascular plants such as the model system *P. patens* possess homologs of these enzymes (Renault *et al.*, 2017a), which suggests a secondary loss of CSE in the respective monocots. CSE belongs to the family of putative MAGLs. MAGLs are found across eukaryotes, and functional analyses in human, yeast and *Arabidopsis* have shown that they possess MAGL

activity (Aschauer *et al.*, 2016; Kim *et al.*, 2016; Labar *et al.*, 2010). In contrast to other MAGLs of *A. thaliana*, AtCSE (MAGL3) was found to exhibit no hydrolytic activity on monoacylglycerols (MAGs) as substrate (Kim *et al.*, 2016)—which applies to other enzymes of *A. thaliana* that belong to the family of MAGLs, too. Indeed, out of the 16 MAGLs that Kim and colleagues (2016) tested, only MAGL6 and 8 showed high activity on MAG as substrate. Given the functional diversity in MAGLs (Kim *et al.*, 2016) and the unequal distribution of caffeoyl-5-*O*-shikimate across embryophytes, functional analyses are required to fully understand how easily MAGLs can lose or gain their MAGL activity. Yet, phylogenetic analyses can pinpoint the diversity of the family across the green lineage.

Here we use phylogenetic analysis to pinpoint the distributions of the diverse MAGL families, including CSE across the green lineage. In total, we recovered all 16 MAGL sequences of *A. thaliana* in the similarity search; using maximum likelihood phylogenetics, we recovered clades for all the 16 MAGLs; some MAGL clades are widely distributed throughout streptophytes, while others appear to have originated in embryophytes, where they have again undergone lineage-specific expansions. Essentially, we recovered two large clades: one restricted to streptophytes, containing homologs of MAGL2, 4 and 13; the other has representation in chlorophytes as well and includes homologs of MAGL1, 3, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15 and 16 (Figure 7).

Focusing on the MAGL2/4/13 clade first, we observe that MAGL13 has representation in angiosperms, gymnosperms and ferns, suggesting its origin to be in the LCA of tracheophytes, while MAGL2 and 4 came from a duplication event before the split between *Arabidopsis* and *Capsella*. However, MAGL2/4 orthologs are present in other species including *P. abies*, suggesting that the common ancestor of seed plants possessed a *MAGL2/4-like* and a *MAGL13* gene. Forming a clade with MAGL2/4/13 are lycophyte, bryophyte and streptophyte algal sequences, which in general branch in an order expected based on their species phylogeny (although within-species duplication events have occurred). This suggests that already at the base of streptophytes a *MAGL2/4/13-like* gene was present.

In the second large clade that includes also chlorophyte sequences, we find the clade containing the CSE/MAGL3 orthologs. This clade includes sequences from both vascular and non-vascular plants, pointing to an origin of CSE in the LCA of land plants. This adds support for a secondary loss in those monocots without a CSE ortholog. Despite the origin of CSE in the common ancestor of land plants and a clear CSE ortholog in *P. patens* (3c19_14430V3.1.p), the substrate of CSE, caffeoyl-5-*O*-shikimate, was not detected in crude extracts of the moss (Renault *et al.*, 2017a). The HCT-based reaction leading to caffeoyl-CoA has been confirmed *in vitro* using moss HCT (Kriegshausen

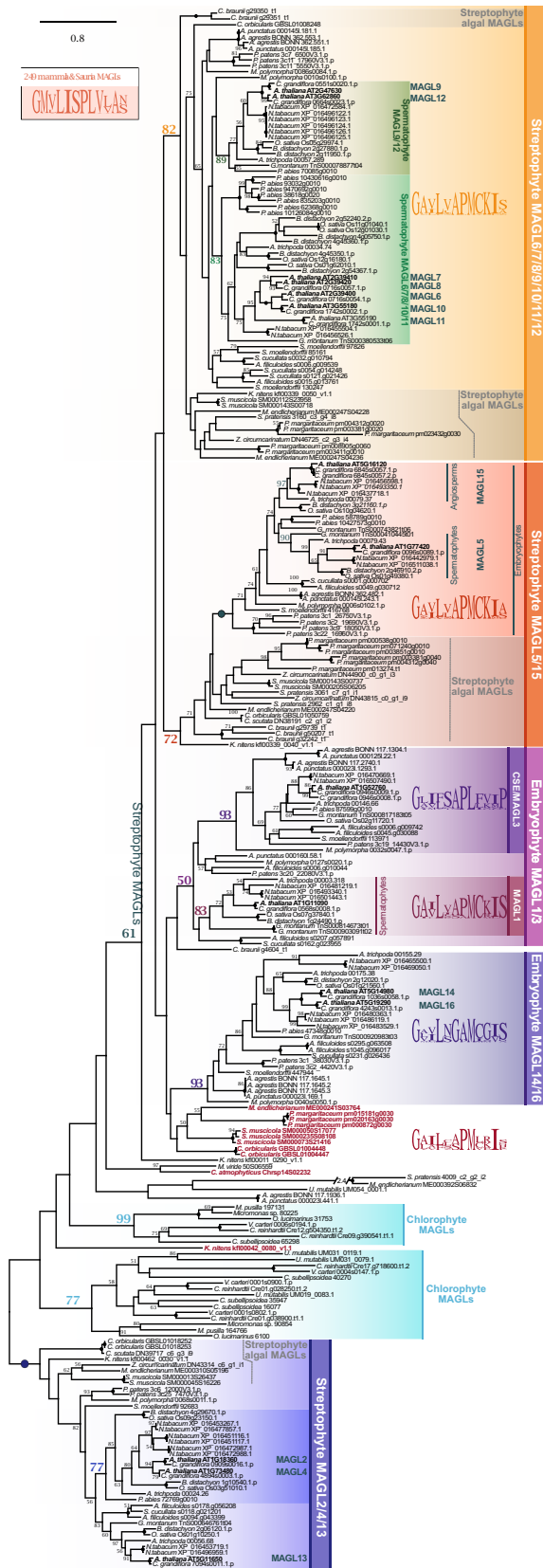


Figure 7. The occurrence of monoacylglycerol lipases (MAGLs) across diverse Streptophyta and a phylogenetic framework for the deep evolutionary roots of caffeoyl-5-O-shikimate esterase (CSE).

MAGL/CSE homologs were sampled from protein data from 15 land plants, seven streptophyte algal and eight chlorophyte algal genomes, as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Coleochaete scutata*, *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *C. orbicularis* (Ju *et al.*, 2015). From all detected homologs, an unrooted maximum likelihood phylogeny was computed using LG+H+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; due to the likely diverse functions, the various clade of MAGL homologs were assigned distinct colors. Purple font highlights those streptophyte algal sequences that share the conserved alpha helix cap domain with CSE. Logos are based on a motif (amino acids 132–141 in MAGL6 and 167–176 in CSE/MAGL3) that is situated in a region likely involved in substrate binding based on the crystal structure of human MAGL.

et al., 2021). Hence, the CSE homologs of *P. patens* may have another function. Indeed, the atypical function of AtCSE, together with the lack of other MAGL family members, to act on MAGs (Kim *et al.*, 2016), suggests that the functional spectrum of the MAGL family is not very limited in land plants.

Members of the MAGL family share several conserved motifs across diverse eukaryotes. One such motif (amino acid positions 132–141 in MAGL6 and 167–176 in CSE/MAGL3) is situated in a region likely involved in substrate binding based on the crystal structure of human MAGL (Labar *et al.*, 2010). Within this motif, a leucine in position number four is found in a diverse set of 249 mammal and Sauria MAGLs investigated here (Figure 7, inset), and most plant MAGLs including MAGL6 and 8. This is followed by another hydrophobic amino acid (isoleucine in mammals/Sauria and valine or leucine in most plant MAGLs). It is striking that exactly these highly conserved amino acids are changed to a phenylalanine and a serine in the Arabidopsis CSE and some homologs from other species. These changes from two very hydrophobic amino acids to an aromatic and a hydrophilic one could be one of the reasons for a change in substrate specificity from a substrate with a hydrophobic acyl chain to a more hydrophilic substrate with aromatic properties. Based on this hypothesis, CSEs would be restricted to some of the members of this clade that, however, stem from across the diversity of land plants.

MAGL1 appears to have originated prior to the split of angio- and gymnosperms, while specific MAGL14 and 16 orthologs likely arose after the split of asterids and rosids, but a MAGL14/16 ortholog was likely present in the LCA of land plants. MAGL15, like MAGL1, originated prior to the split of gymno- and angiosperms, and MAGL5 appeared to come from a duplication of MAGL15 later on possibly in the ancestors of dicots. Interestingly, a *MAGL5/15-like* sequence

Figure 8. Low resolution on the complex evolutionary history of caffeate *O*-methyltransferase (COMT).

We explored the diversity of methyltransferases by screening for sequences homologous to Arabidopsis COMT/OMT1 across genome data from 15 land plants, seven streptophyte algae and five chlorophytes; additionally, we included sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *Coleochaete orbicularis* (Ju *et al.*, 2015). From all detected homologs, an unrooted maximum likelihood phylogeny of 226 sequences was computed using LG+I+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; homologs with known COMT function were assigned the warmest and brightest reddish colors. Blue font highlights streptophyte algal sequences; bold font pinpoints those that recovered land plant COMT as closest structural analogs in I-TASSER-based modeling. On the right we show residues important for substrate binding and function of canonical COMT, as reported by Louie *et al.* (2010).

was already encoded in the genome of the ancestor of streptophytes. The same is true for a *MAGL6/7/8/9/10/11/12-like* gene, which similar to the *MAGL5/15-like* genes shows independent paths of radiation in streptophyte algae and land plants. *MAGL6,7,8,10* and *11* are only present in the here included Brassicaceae, while a *MAGL6/7/8/10/11-like* gene was already present in the common ancestor of spermatophytes. The same evolutionary history describes the scenario under which *MAGL9* and *12* originated.

Finding the *MAGL6/7/8/10/11* subclade specifically expanded in seed plants is noteworthy. At least *AtMAGL8* localizes to lipid droplets (Kim *et al.*, 2016), which are structures found in various photosynthetic eukaryotes and are well known from seeds. Thus, expansion of this clade might be a read-out of spermatophyte-specific additions to the ancient set of proteins relevant to lipid droplet formation and function (de Vries and Ischebeck, 2020).

All in all, MAGLs have undergone an early radiation in streptophytes. Given that even the Arabidopsis MAGLs without detectable activity on MAG (Kim *et al.*, 2016) do not form a monophylum, it is conceivable that subfunctionalization of members of the MAGL family occurred multiple times independently. This may likewise be true for all independent expansions of MAGL-encoding genes observed in any other species included here. The versatility in functional evolution of MAGLs makes it difficult to make robust predictions of putative MAGL functions.

Caffeate *O*-methyltransferase (COMT): convergence and complexity

In angiosperms, ferulate 5-hydroxylase (F5H) and COMT carry out important catalytic steps along the route from *p*-coumaroyl-CoA to S-lignin. COMT catalyzes the methylation of caffeic acid or 5-hydroxyferulic acid, the product formed by F5H. Like C3H and C4H, F5H belongs to the large CYP450 clan 71 (Nelson and Werck-Reichhart, 2011). The function of F5H evolved at least twice in P450 enzymes, once in the ancestor of angiosperms and once in the ancestor of lycophytes (Weng and Chapple, 2010; Weng *et al.*, 2008).

For angiosperm F5H, no clear putative orthologs were found outside of flowering plants, and likewise no clear orthologs were found for the lycophyte F5H (i.e. 'SmF5H'), which forms a separate clade from the angiosperm F5H sequences (Figure S11). This corroborates previous results (de Vries *et al.*, 2017), and is in agreement with the

hypothesis that F5H function evolved at least twice in the evolution of land plants (Weng *et al.*, 2008). Additionally, the average pairwise identity of the F5H homologs was low (18.5%)—hampering robust phylogenetic analyses. We thus did not further delve into the evolution of F5H. COMT, however, caught our attention.

The lycophyte *S. moellendorffii* not only uses a genetically distant F5H enzyme; the same appears to be true for COMT (Weng *et al.*, 2011). This highlights a promiscuity for substrate specificity and activity in P450 enzymes that is yet to be discovered, and mere orthology analyses can only go so far as to discover putative candidates. Using phylogenetics, we explored the diversity of methyltransferases by screening for sequences homologous to COMT/OMT1 of *A. thaliana* across our phylo diverse dataset. This approach identified not only clear orthologs, but can also serve as a backbone to map relevant amino acid substitutions facilitating in functional convergence in this group of enzymes, and by that may highlight possible candidates for *in vivo* and *in vitro* studies.

For most clades of land-plant methyltransferases, based on the here recovered topology, predicting a putative function was not straightforward. This applied even more so to the homologs of COMT/OMT1 found in chlorophyte and streptophyte algae. We recovered a clade of methyltransferases that included chlorophyte and streptophyte algae as well as diverse land plant sequences (coined 'Chloroplastida OMT' in Figure 8); among these clustered *A. thaliana* proteins such as COMT, indole glucosinolate methyltransferases (IGMT) and nicotinate *N*-methyltransferase (NANMT; Li *et al.*, 2017)—hence different methyltransferases that act on a range of aromatic compounds. What this means for the presence of a putative COMT in algae is obscure. However, it corroborates the previously observed patchy detection of COMT across the green lineage based on reciprocal BLASTp searches (de Vries *et al.*, 2017). Clear orthologs of *AtCOMT* were only detected for a few angiosperms, notably not including any monocot sequence that we used in our dataset. Of all methyltransferases in our dataset, only NANMT formed a clade of clear orthologs that included more than one major lineage of land plants by encasing sequences from angiosperms and *P. abies* (bootstrap support 87). All other orthogroups appear, like COMT, to be restricted to only a few of the included angiosperm lineages.



Figure 9. A phylogenetic framework for the evolutionary origin of caffeoyl-CoA *O*-methyltransferase (CCoAOMTs) in Phragmoplastophyta. CCoAOMT homologs were sampled from protein data from 16 land plants, seven streptophyte algal and eight chlorophyte algal genomes, as well as sequences found in the transcriptomes of *Spirogyra pratensis* (de Vries *et al.*, 2020), *Coelaechaete scutata*, *Zygnema circumcarinatum* (de Vries *et al.*, 2018) and *C. orbicularis* (Ju *et al.*, 2015). From all detected homologs, an unrooted maximum likelihood phylogeny of 138 sequences was computed using LG+G4 as model for protein evolution (chosen according to BIC). One-hundred bootstrap replicates were computed; only bootstrap values ≥ 50 are shown, and bootstrap values of 100 are depicted by a filled dot. Colored font and dots correspond to the support recovered for the higher-order clades labeled on the right of the phylogeny; the large clade of the diverse *S*-adenosyl-L-methionine-dependent methyltransferases was assigned the color red. The alignment on the right shows functionally characterized sites involved in substrate, ion and co-factor recognition of CCoAOMT (Ferrer *et al.*, 2005).

The lycophyte *S. moellendorffii* has a COMT that is distantly related to COMT of angiosperms. It appears to have acquired its COMT activity through convergent evolution, and was coined *Sm*COMT (Weng *et al.*, 2011). In agreement with this, *Sm*COMT did not cluster with the *At*OMT1 sequence in our analyses. Instead, it forms its own (weakly supported) clade with only one other sequence from *S. moellendorffii* (bootstrap support 63). The other *Sm*COMT-like sequences (described in Weng *et al.*, 2011) were distributed over the phylogeny and appear to be specific to *S. moellendorffii*. Nonetheless, this pattern highlights a certain versatility in the evolutionary history of substrate specificity of *O*-methyltransferases in land plants. Indeed, despite the lack of a clear ortholog to *Arabidopsis* COMT, evidence for the presence of *O*-methyltransferases with COMT-like activity in pine wood has been brought forward (Wagner *et al.*, 2015). This appears to be only logical, noting the large lineage-specific expansions in the larger clade of *O*-methyltransferases that encompasses all *O*-methyltransferases from *A. thaliana*—that is COMT, IGMTs, NANMT and *N*-acetylserotonin *O*-methyltransferase (ASMT, bootstrap-support 99; for more on this enzyme, see Byeon *et al.*, 2016; Tan *et al.*, 2012). Within this clade fall also algal sequences from chlorophytes and streptophyte algae. Their position within the clade is undetermined due to low bootstrap support. These sequences appear highly divergent, many of them cluster with rather long branches. Yet, some of the sequences from our previous analysis found a reciprocal BLASTp hit to *At*COMT, including *K. nitens* 00158_0100v1.1 that clusters in a fully supported clade of *Klebsormidium* paralogs; these sequences are promising candidates to explore COMT activity. Indeed, when we modeled the tertiary structure of *K. nitens* 00158_0100v1.1 and *M. endlicherianum* ME000591S08520 using I-TASSER (Zhang, 2008), we recovered *Medicago sativa* and *Lolium perenne* COMT as its closest structural analogs (1KYZ; Zubieta *et al.*, 2002; 3P9C; Louie *et al.*, 2010; TM-scores 0.865 and 0.956, respectively).

Like the land plant COMTs, also the algal COMT-like sequences appear to have undergone independent radiations in this large gene family. Given the observed convergent evolution of COMT activity in *S. moellendorffii*, the question of whether there is COMT activity across Streptophyta remains wide open.

To gain some insight into whether COMT activity can be expected from other streptophyte lineages, we

investigated the conservation of residues relevant for the function of COMT, including those that form the substrate-binding pocket (Figure 8). The functional residues were identified from COMT of *L. perenne* (Louie *et al.*, 2010). Across our phylogeny, these residues differ between the clades of canonical ASMT, NANMT, COMT and IGMT, while they are conserved within them (Figure 8). The binding pocket of *At*OMT1 and its orthologs consist of the amino acid pattern MSNGGG, whereas the pattern for the residues important for the function of the enzyme is HDE. While HDE appears conserved across the majority of sequences analyzed here, independent of the specific function of the enzyme (e.g. ASMT, IGMTs and COMT all have the pattern HDE), the binding pocket is highly variable among the functionally characterized enzymes. This suggests that the reaction-determining residues are those that form the binding pocket and not those that are catalytically important. This seems logical given that all these enzymes catalyze similar types of reactions. The triple G in the binding pocket is also far more conserved across the entire phylogeny, with only a few exceptions occurring, while the first three residues are highly variable. Indeed, the COMT-specific MSN motif is not present in the functionally characterized COMT from *S. moellendorffii*, rather it is MTN, which, however, is a change between similar amino acids (Ser to Thr). Apart from the canonical COMT orthologs and *Sm*COMT sequence, no other sequences from any other lineage encode the binding pocket pattern M(S/T)NGGG, suggesting that none has a canonical preference for binding 5-hydroxyconiferinaldehyde. Yet, several homologs—including those of streptophyte algae—would have the ability to catalyze the reaction based on the conservation of the residue pattern HDE. What is more, all the sequences that could not be properly identified as orthologs to ASMT, NANMT, COMT and IGMT (with the exception of *Sm*COMT) show no similarity in their first three residues of the binding pocket to either of these enzyme families. This would suggest that the enzymes from most streptophyte lineages included in this analysis use different substrates than those functionally characterized in *A. thaliana*, again underscoring the versatility of this family.

Caffeoyl-CoA *O*-methyltransferase (CCoAOMT)-like sequences emerged in Phragmoplastophyta

Within the phenylpropanoid pathway, CCoAOMT and most of its homologs are the enzymes that catalyze the first

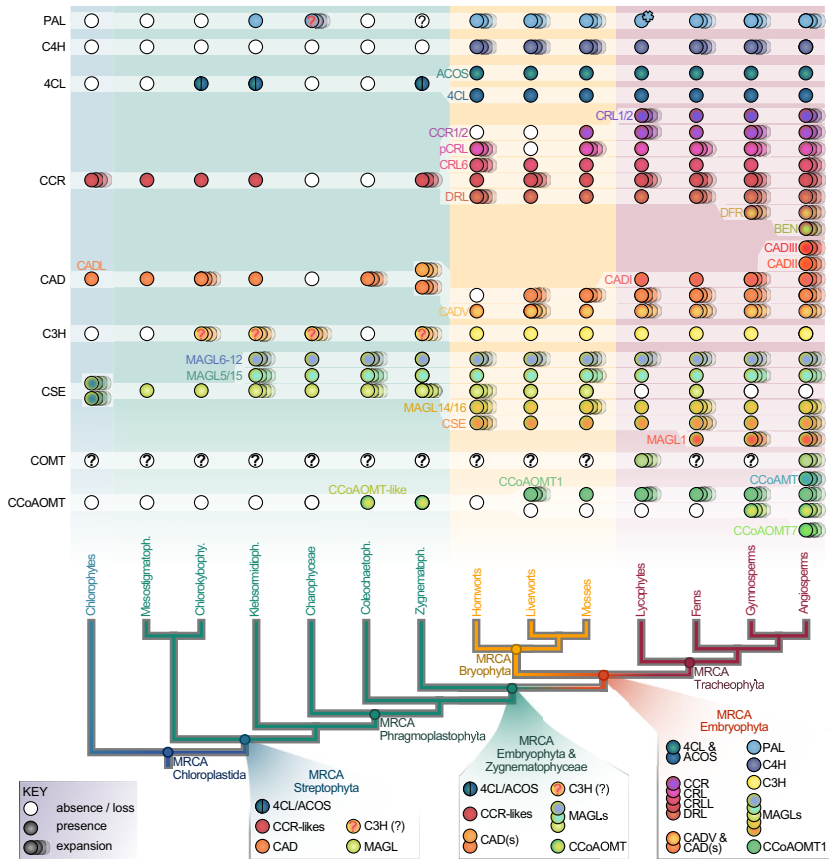


Figure 10. A summary of the proposed evolutionary trajectory of key enzymes in the phenylpropanoid pathway across the green lineage. At the bottom is a cladogram of the green lineage. The most recent common ancestors (MRCA) of Chloroplastida, Streptophyta, Phragmoplastophyta, Bryophyta, Embryophyta and Tracheophyta are indicated at their respective nodes. On top of the cladogram is the proposed evolutionary trajectory of the enzyme families phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate-CoA ligase (4CL), cinnamoyl-CoA reductase (CCR), cinnamyl alcohol dehydrogenase (CAD), 4-coumarate 3-hydroxylase (C3H), caffeoyl-5-*O*-shikimate esterase (CSE)/monoacylglycerol lipase (MAGL), caffeate *O*-methyltransferase (COMT), and caffeoyl-CoA *O*-methyltransferase (CCoAOMT). The names of the enzyme families are indicated on the left of the trajectory. The enzyme (sub-)families present in a specific common ancestor have been plotted onto the respective nodes of the cladogram below the evolutionary scenario of the enzyme families involved in the phenylpropanoid pathway and lignin biosynthesis. White dots indicate absence/loss of a gene family, one dot indicates the presence of one representative of the gene family and several dots indicate an expansion (two or more members of the gene family) in at least one species of the represented lineages in the cladogram. Colors are chosen to distinguish different enzyme families and subfamilies. Question marks label sequences of ambiguous affiliation.

committed step to many of at least two types of lignin (S- and G-lignin). These enzymes methylate caffeoyl-CoA and thus give rise to feruloyl-CoA (Do *et al.*, 2007; Martz *et al.*, 1998; Vanholme *et al.*, 2012; Ye *et al.*, 1994; Ye and Varner, 1995); in the past, it was also proposed that after the conversion to 5'-hydroxy-feruloyl-CoA, CCoAOMT can methylate this compound to produce Sinapoyl-CoA (Ferrer *et al.*, 2005; Maury *et al.*, 1999). One of the paralogs that exist in *A. thaliana* [tapetum-specific *O*-methyltransferase (TSM1), AT1G67990], however, shows activity towards a coniferyl derivative that is formed at N^{10} by F5H starting from N^1, N^5, N^{10} -tris-(hydroxyferuloyl) spermidine (Fellenberg *et al.*, 2008); TSM1 catalyzes the production of N^1, N^5 -bis-(hydroxyferuloyl)- N^{10} -synapoylspermidine (Fellenberg *et al.*, 2008). Thus, CCoAOMTs appear to be versatile in their substrate specificity and can act on different steps in the phenylpropanoid pathway.

Here we used phylogenetics to disentangle the distribution of these enzymes across the green lineage. A duplication gave rise to the genes encoding the functionally divergent enzymes *AtCCoAMT* and *AtTSM1* (bootstrap support 77; Figure 9). These two methyltransferases are embedded in a larger clade containing the other CCoAOMT enzymes *AtCCoAMT1*, *AtCCoAMT7* and *AtCCoAMT-like* (AT1G24735). The latter appears to be

specific to the Brassicaceae included in this dataset, while homologs of *AtCCoAMT7* occur across dicots and were detected in *A. trichopoda*, but were absent from the included monocots. Only *AtCCoAMT1* had a wider distribution. Its cluster (bootstrap support 79) contains angiosperms, gymnosperms, lycophytes, ferns and bryophytes, excluding the sequenced hornworts from the genus *Anthoceros* (Figure 9). Assuming a monophyly of Bryophyta (Puttick *et al.*, 2018), this suggests a loss of CCoAOMT1 in at least the sequenced *Anthoceros* species, and that CCoAOMT1 was present in the common ancestor of land plants. Lineage-specific duplications of CCoAOMT1 appear to have happened, indicated by the expansions seen in tobacco, spruce, the lycophyte *S. moellendorffii* and the water fern *A. filiculoides*. The expanded repertoire of sequences in monocots and *Gnetum* indicate additional lineage-specific duplications outside of the CCoAOMT clade. The case of TSM1 suggests that neofunctionalization can easily occur within this type of methyltransferase. We noted that the residues involved in substrate binding (Ferrer *et al.*, 2005) are identical in *AtCCoAMT* and *AtTSM1* (Figure 9). A possible explanation might be that the make-up of the binding pocket allows for a certain versatility in substrates. Given these observations, the paralogs within this and the other CCoAOMT

clades cannot be assumed to hold the function of CCoAOMTs. Likewise, it cannot be ruled out that their LCA may have had this function. As sister to the methyltransferase clade, including the CCoAOMT homologs lies a cluster of genes encoding putative candidates for streptophyte algal CCoAOMTs. These were limited to representatives of the two streptophyte algal lineages closest to land plants: the Coleochaetophyceae *Coleochaete scutata* and *C. orbicularis*, as well as the Zygnematophyceae *Spirogløea muscicola*. These algal sequences have the same domain structure as the majority of all S-adenosyl-L-methionine (SAM)-dependent methyltransferases included in the phylogeny (Figure S12). Only single sequences, scattered across the phylogeny and diversity of species included here, vary in their domain structure, showing a loss of a domain, a gain of an additional domain, or both. The analyses of the specific functional residues gave more insights into the streptophyte algal sequences within the clade of SAM-dependent methyltransferases in the peripheral routes of the phenylpropanoid pathway. These algal sequences maintain the residues involved in ion and cofactor binding, but differ strongly in the substrate binding site (Ferrer *et al.*, 2005; Figure 9).

Altogether, it appears that the family of CCoAOMT-like proteins has its origin in Phragmoplastophyta. Clarifying the function of the putative CCoAOMT-like enzymes in Coleochaetophyceae and Zygnematophyceae has the potential to shed light on a synapomorphy with physiological relevance.

Conclusion

All genes for enzymes that act in early steps in the chassis of the phenylpropanoid pathway investigated here (Figure 1) can be traced back to the LCA of land plants with the exception of COMT (Figure 10); most of these can even be traced back to some ancestor that land plants shared with streptophyte algae. While most of our knowledge on how these genes work comes from angiosperms, this does not capture the sequence diversity in enzymes—and it underpins the versatility in producing specialized metabolites.

Our data pinpoint that most of the enzymes have undergone pronounced lineage-specific expansions. A lineage-specific expansion is palpable even despite the fact that sampling of sequences across the Streptophyta is still strongly biased towards seed plants. These data offer a framework for pinpointing those candidate genes/enzymes that are bound to shed light on the evolution of key enzymatic steps—and currently unknown ones. Such work is exemplified by studies on *Selaginella* or bryophyte model systems such as *P. patens*. Characterizing enzymes that are even more divergent from what we know from angiosperms should yield surprising insights and unexplored routes in this bountiful pathway.

EXPERIMENTAL PROCEDURES

Dataset of protein sequences and screening for homologs

We downloaded protein data from: (a) genomes of 16 land plants: *A. agrestis* as well as *Anthoceros punctatus* (Li *et al.*, 2020), *A. tri-chopoda* (Amborella Genome Project, 2013), *A. thaliana* (Lamesch *et al.*, 2011), *A. filiculoides* (Li *et al.*, 2018), *B. distachyon* (The International Brachypodium Initiative, 2010), *C. grandiflora* (Slotte *et al.*, 2013), *G. montanum* (Wan *et al.*, 2018), *M. polymorpha* (Bowman *et al.*, 2017), *Nicotiana tabacum* (Sierro *et al.*, 2014), *Oryza sativa* (Ouyang *et al.*, 2007), *P. abies* (Nystedt *et al.*, 2013), *P. patens* (Lang *et al.*, 2018), *Salvinia cucullata* (Li *et al.*, 2018), *S. moellendorffii* (Banks *et al.*, 2011) and *Theobroma cacao* (Argout *et al.*, 2011); (b) the genomes of seven streptophyte algae: *C. atmophyticus* (Wang *et al.*, 2020), *C. braunii* (Nishiyama *et al.*, 2018), *K. nitens* (Hori *et al.*, 2014), *M. endlicherianum* (Cheng *et al.*, 2019), *Mesostigma viride* (Wang *et al.*, 2020), *Penium margaritaceum* (Jiao *et al.*, 2020), *S. muscicola* (Cheng *et al.*, 2019); (c) the genomes of eight chlorophytes: *Bathycoccus prasinos* (Moreau *et al.*, 2012), *Chlamydomonas reinhardtii* (Merchant *et al.*, 2007), *Coccomyxa subellipsoidea* (Blanc *et al.*, 2012), *Micromonas pusilla*, *Micromonas* sp. (Worden *et al.*, 2009), *Ostreococcus lucimarinus* (Palenik *et al.*, 2007), *Ulva mutabilis* (De Clerck *et al.*, 2018), *Volvox carteri* (Prochnik *et al.*, 2010). Additionally, we included sequences found in the transcriptomes of *S. pratensis* (de Vries *et al.*, 2020), *Coleochaete scutata* and *Z. circumcarinatum* (de Vries *et al.*, 2018) and *C. orbicularis* (Ju *et al.*, 2015).

For each of the protein families investigated, the representative *A. thaliana* protein was used as a query sequence for a BLASTp against this dataset. Initially, we considered all homologs recovered at an *e*-value cutoff level of 10^{-7} . However, due to the large size of the protein families (i.e. high number of well-supported homologs obtained), refinement of the datasets was carried out as described in the individual sections for these enzymes in the Results and Discussion section.

Alignments, phylogenetic analysis and primary sequence analysis

Using the detected homologs for a given enzyme family, we generated alignments using MAFFT v7.453 (Katoh and Standley, 2013) with a L-INS-I approach. Alignments were cropped, if necessary, to retain conserved domains that were alignable for all homologs; alignments are provided in Supplemental Data S1–S11. We computed maximum likelihood phylogenies using IQ-TREE multicore version 1.5.5 (Nguyen *et al.*, 2015), with 100 bootstrap replicates. To determine the best model, we used ModelFinder (Kalyaanamoorthy *et al.*, 2017) and picked the best models based on the Bayesian Information Criterion. The best models were: LG+G4 (Le and Gascuel, 2008) for 4CL, CCR, CCoAOMT; LG+I+G4 for PAL, CAD, MAGL/CSE, COMT, and for the preliminary phylogeny of 4CL; LG+F + I+G4 for C4H, F5H, and C3H; WAG+F+G4 (Whelan and Goldman, 2001) for HCT.

Protein structure prediction was carried out using the sequences as input in the online Iterative Threading ASSEmbly Refinement (I-TASSER; Yang *et al.*, 2015; Zhang, 2008). Functional residue analyses were based on published structural analyses (Ferrer *et al.*, 2005; Hu *et al.*, 2010; Pan *et al.*, 2014; Rupasinghe *et al.*, 2003; Youn *et al.*, 2006), and alignments were viewed with SeaView v.4 (Gouy *et al.*, 2010) and plotted with ETE3 (Huerta-Cepas *et al.*, 2016).

Protein domains of the enzyme families 4CL, CCR, C3H, CAD and CCoAOMT were predicted using InterProScan version 5.47-

82.0 (Jones *et al.*, 2014). Presence/absence heatmaps were generated and projected onto phylogenies (visualized using iTOL v6; Letunic and Bork, 2019).

DATA AVAILABILITY STATEMENT

All sequence data used for this article are publicly available, based on published genomes and transcriptomes, and cited in the Experimental Procedures section. All other data can be found within the manuscript and its supporting materials; this includes the alignments, which have additionally been deposited at Zenodo under <https://doi.org/10.5281/zenodo.5011621>.

ACKNOWLEDGEMENTS

J.M.R.F.-J. is grateful for being supported by the Ph.D. program 'Microbiology and Biochemistry' within the framework of the 'Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences' (GGNB) at the University of Goettingen; A.D.A. is grateful for being supported through the International Max Planck Research School (IMPRS) for Genome Science. J.d.V. thanks the European Research Council for funding under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 852725; ERC-StG 'TerreStriAL'). M.P., I.F. and J.d.V. are grateful for support through the German Research Foundation (DFG) within the framework of the Priority Programme 'MAdLand – Molecular Adaptation to Land: Plant Evolution to Change' (SPP 2237; VR 132/4-1; PE 360/37-1; FE 446/14-1). Open Access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

JdV designed the study. SdV, JMRF-J, II, ADA and JdV performed bioinformatic analyses and designed the figures. SdV, II, TI, KF, INA, MP, IF and JdV interpreted the results. SdV and JdV wrote the initial draft with comments from II, TI, MP and IF. All authors contributed to writing the final manuscript. All authors discussed the results.

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Predicted structures for PAL homologs.

Figure S2. Phylogeny of C4H homologs.

Figure S3. Phylogeny of 4CLs/AAs.

Figure S4. Domains predicted for protein sequences in the phylogeny of 4CL/AAs homologs.

Figure S5. Domains predicted for protein sequences in the phylogeny of CCR homologs.

Figure S6. Key residues of protein sequences in the phylogeny of CCR homologs.

Figure S7. Domains predicted for protein sequences in the phylogeny of CAD homologs.

Figure S8. Phylogeny of HCT homologs.

Figure S9. Key residues of protein sequences in the phylogeny of C3H homologs.

Figure S10. Domains predicted for protein sequences in the phylogeny of C3H homologs.

Figure S11. Phylogeny of F5H homologs.

Figure S12. Domains predicted for protein sequences in the phylogeny of CCoAOMT homologs.

Figure S13. Uncollapsed phylogeny of C3H homologs.

Table S1. Patristic distances of the phylogeny shown in Figure S2

Data S1. Multiple sequence alignment (FASTA format) used for the phylogeny of PAL homologs shown in Figure 2.

Data S2. Multiple sequence alignment (FASTA format) used for the phylogeny of 4CL homologs shown in Figure 3.

Data S3. Multiple sequence alignment (FASTA format) used for the phylogeny of CCR homologs shown in Figure 4.

Data S4. Multiple sequence alignment (FASTA format) used for the phylogeny of CAD homologs shown in Figure 5.

Data S5. Multiple sequence alignment (FASTA format) used for the phylogeny of C3H homologs shown in Figure 6.

Data S6. Multiple sequence alignment (FASTA format) used for the phylogeny of CSE/MAGL homologs shown in Figure 7.

Data S7. Multiple sequence alignment (FASTA format) used for the phylogeny of COMT homologs shown in Figure 8.

Data S8. Multiple sequence alignment (FASTA format) used for the phylogeny of CCoAOMT homologs shown in Figure 9.

Data S9. Multiple sequence alignment (FASTA format) used for the phylogeny of C4H homologs shown in Figure S2.

Data S10. Multiple sequence alignment (FASTA format) used for the phylogeny of HCT homologs shown in Figure S8.

Data S11. Multiple sequence alignment (FASTA format) used for the phylogeny of F5H homologs shown in Figure S11.

REFERENCES

- Amborella Genome Project** (2013) The *Amborella* genome and the evolution of flowering plants. *Science*, **342**, 1241089.
- Argout, X., Salse, J., Aury, J.-M., Guiltinan, M.J., Droc, G., Gouzy, J. et al.** (2011) The genome of *Theobroma cacao*. *Nature Genetics*, **43**, 101–108.
- Aschauer, P., Rengachari, S., Lichtenegger, J., Schittmayer, M., Das, K. M., Mayer, N. et al.** (2016) Crystal structure of the *Saccharomyces cerevisiae* monoglyceride lipase Yju3p. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, **1861**, 462–470. <https://doi.org/10.1016/j.bbalip.2016.02.005>
- Baedecker, M. & Schulz, G.E.** (2002) Structures of two histidine ammonia-lyase modifications and implications for the catalytic mechanism. *European Journal of Biochemistry*, **269**, 1790–1797.
- Banks, J.a., Nishiyama, T., Hasebe, M., Bowman, J.I., Gribskov, M., dePamphilis, C. et al.** (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science*, **332**, 960–963.
- Barakat, A., Bagniewska-Zadworna, A., Choi, A., Plakkat, U., DiLoreto, D.S., Yellanki, P. et al.** (2009) The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC Plant Biology*, **9**, 26.
- Barakate, A., Stephens, J., Goldie, A., Hunter, W.N., Marshall, D., Hancock, R.D. et al.** (2011) Syringyl lignin is unaltered by severe sinapyl alcohol dehydrogenase suppression in tobacco. *The Plant Cell*, **23**, 4492–4506.
- Barros, J. & Dixon, R.A.** (2020) Plant phenylalanine/tyrosine ammonia-lyases. *Trends in Plant Science*, **25**, 66–79.
- Barros, J., Serrani-Yarce, J.C., Chen, F., Baxter, D., Venables, B.J. & Dixon, R.A.** (2016) Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants*, **2**, 16050.
- Bednarek, P., Schneider, B., Svatoš, A., Oldham, N.J. & Hahlbrock, K.** (2005) Structural complexity, differential response to infection, and tissue specificity of indolic and phenylpropanoid secondary metabolism in *Arabidopsis* roots. *Plant Physiology*, **138**, 1058–1070.

- Berens, M.L., Berry, H.M., Mine, A., Argueso, C.T. & Tsuda, K. (2017) Evolution of hormone signaling networks in plant defense. *Annual Review of Phytopathology*, **55**, 401–425.
- Berland, H., Albert, N.W., Stavland, A., Jordheim, M., McGhie, T.K., Zhou, Y. *et al.* (2019) Auronidins are a previously unreported class of flavonoid pigments that challenges when anthocyanin biosynthesis evolved in plants. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 20232–20239.
- Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D.D. *et al.* (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biology*, **13**, R39.
- Blázquez, M.A., Nelson, D.C. & Weijers, D. (2020) Evolution of plant hormone response pathways. *Annual Review of Plant Biology*, **71**, 327–353.
- Booij-James, I.S., Dube, S.K., Jansen, M.A.K., Edelman, M. & Mattoo, A.K. (2000) Ultraviolet-B radiation impacts light-mediated turnover of the photosystem II reaction center heterodimer in *Arabidopsis* mutants altered in phenolic metabolism. *Plant Physiology*, **124**, 1275–1284.
- Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K. *et al.* (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell*, **171**, 287–304.
- Byeon, Y., Lee, H.J., Lee, H.Y. & Back, K. (2016) Cloning and functional characterization of the *Arabidopsis* N-acetylserotonin O-methyltransferase responsible for melatonin synthesis. *Journal of Pineal Research*, **60**, 65–73.
- Carella, P., Gogleva, A., Hoey, D.J., Bridgen, A.J., Stolze, S.C., Nakagami, H. *et al.* (2019) Conserved biochemical defenses underpin host responses to oomycete infection in an early-divergent land plant lineage. *Current Biology*, **29**, 2282–2294.e5.
- Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T. *et al.* (2019) Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell*, **179**, 1057–1067.e14.
- Chezem, W.R., Memon, A., Li, F.-S., Weng, J.-K. & Clay, N.K. (2017) SG2-type R2R3-MYB transcription factor MYB15 controls defense-induced lignification and basal immunity in *Arabidopsis*. *The Plant Cell*, **29**, 1907–1926.
- Clayton, W.A., Albert, N.W., Thrimawithana, A.H., McGhie, T.K., Derolles, S.C., Schwinn, K.E. *et al.* (2018) UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *The Plant Journal*, **96**, 503–517.
- Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.-W., Cardenas, C.L., Cochran, F.C. *et al.* (2005) Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *Phytochemistry*, **66**, 2072–2091.
- Danielsson, M., Lundén, K., Elfstrand, M., Hu, J., Zhao, T., Arnerup, J. *et al.* (2011) Chemical and transcriptional responses of Norway spruce genotypes with different susceptibility to *Heterobasidion* spp. infection. *BMC Plant Biology*, **11**, 154.
- D'Auria, J.C. (2006) Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology*, **9**, 331–340.
- de Azevedo Souza, C., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M. *et al.* (2009) A novel fatty acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. *The Plant Cell*, **21**, 507–525.
- De Clerck, O., Kao, S.-M., Bogaert, K.A., Blomme, J., Foflonker, F., Kwantes, M. *et al.* (2018) Insights into the evolution of multicellularity from the sea lettuce genome. *Current Biology*, **28**, 2921–2933.e2925.
- de Vries, J. & Archibald, J.M. (2018) Plant evolution: landmarks on the path to terrestrial life. *New Phytologist*, **217**, 1428–1434.
- de Vries, J., Curtis, B.A., Gould, S.B. & Archibald, J.M. (2018) Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E3471–E3480.
- de Vries, J., de Vries, S., Curtis, B.A., Zhou, H., Penny, S., Feussner, K. *et al.* (2020) Heat stress response in the closest algal relatives of land plants reveals conserved stress signalling circuits. *The Plant Journal*, **103**, 1025–1048.
- de Vries, J., de Vries, S., Slamovits, C.H., Rose, L.E. & Archibald, J.M. (2017) How embryophytic is the biosynthesis of phenylpropanoids and their derivatives in streptophyte algae? *Plant and Cell Physiology*, **58**, 934–945.
- de Vries, J. & Ischebeck, T. (2020) Ties between stress and lipid droplets pre-date seeds. *Trends in Plant Science*, **12**, 1203–1214.
- de Vries, S., Herrfurth, C., Li, F.-W., Feussner, I. & de Vries, J. (2021) An ancient route towards salicylic acid and its implications for the perpetual *Trichormus-Azolla* symbiosis. *bioRxiv* preprint <https://doi.org/10.1101/2021.03.12.435107>
- Delwiche, C.F., Graham, L.E. & Thomson, N. (1989) Lignin-like compounds and sporopollenin in *Coleochaete*, an algal model for land plant ancestry. *Science*, **245**, 399–401.
- Devic, M., Guillemot, J., Debeaujon, I., Bechtold, N., Bensaude, E., Koornneef, M. *et al.* (1999) The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *The Plant Journal*, **19**, 387–398.
- Dixon, R.A., Achnine, L., Kota, P., Liu, C.-J., Srinivasa Reddy, M.S. & Wang, L. (2002) The phenylpropanoid pathway and plant defence—a genomics perspective. *Molecular Plant Pathology*, **3**, 371–390.
- Dixon, R.A. & Paiva, N.L. (1995) Stress-induced phenylpropanoid metabolism. *The Plant Cell*, **7**, 1085–1097.
- Do, C.-T., Pollet, B., Thévenin, J., Sibout, R., Denoue, D., Barrière, Y. *et al.* (2007) Both caffeoyl coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in *Arabidopsis*. *Planta*, **226**, 1117–1129.
- Ehltling, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. & Kombrink, E. (1999) Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *The Plant Journal*, **19**, 9–20.
- Emiliani, G., Fondi, M., Fani, R. & Gribaldo, S. (2009) A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. *Biology Direct*, **4**, 7.
- Eudes, A., Pereira, J.H., Yogiswara, S., Wang, G., Benites, V.T., Baidoo, E.E.K. *et al.* (2016) Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase to reduce lignin. *Plant and Cell Physiology*, **57**, 568–579.
- Fellenberg, C., Milkowski, C., Hause, B., Lange, P.-R., Böttcher, C., Schmidt, J. *et al.* (2008) Tapetum-specific location of a cation-dependent O-methyltransferase in *Arabidopsis thaliana*. *The Plant Journal*, **56**, 132–145.
- Ferrer, J.L., Zubieta, C., Dixon, R.A. & Noel, J.P. (2005) Crystal structures of alfalfa caffeoyl coenzyme A 3-O-methyltransferase. *Plant Physiology*, **137**, 1009–1017.
- Franks, N.P., Jenkins, A., Conti, E., Lieb, W.R. & Brick, P. (1998) Structural basis for the inhibition of firefly luciferase by a general anesthetic. *Biochemical Journal*, **75**, 2205–2211.
- Fürst-Jansen, J.M.R., de Vries, S. & de Vries, J. (2020) Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany*, **11**, 3254–3269.
- Goiris, K., Muylaert, K., Voorspoels, S., Noten, B., De Paepe, D., Baart, G.J.E. *et al.* (2014) Detection of flavonoids in microalgae from different evolutionary lineages. *Journal of Phycology*, **50**, 483–492.
- Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224.
- Grienenberger, E., Kim, S.S., Lallemand, B., Geoffroy, P., Heintz, D., de Azevedo Souza, C. *et al.* (2010) Analysis of TETRAKETIDE α -PYRONE REDUCTASE function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *The Plant Cell*, **22**, 4067–4083.
- Gross, G.G., Stöckigt, J., Mansell, R.L. & Zenk, M.H. (1973) Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate: CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. *FEBS Letters*, **31**, 283–286.
- Güngör, E., Brouwer, P., Dijkhuizen, L.W., Shaffar, D.C., Nierop, K.G.J., de Vos, R.C.H. *et al.* (2021) *Azolla* ferns testify: seed plants and ferns share a common ancestor for leucoanthocyanidin reductase enzymes. *New Phytologist*, **229**, 1118–1132.
- Guo, D.-M., Ran, J.-H. & Wang, X.-Q. (2010) Evolution of the Cinnamyl/Sinapyl Alcohol Dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of bona fide CAD. *Journal of Molecular Evolution*, **71**, 202–218.
- Ha, C.M., Escamilla-Trevino, L., Yance, J.C.S., Kim, H., Ralph, J., Chen, F. *et al.* (2016) An essential role of caffeoyl shikimate esterase in monoglucosyl biosynthesis in *Medicago truncatula*. *The Plant Journal*, **86**, 363–375.
- Hamberger, B., Ellis, M., Friedmann, M., de Azevedo Souza, C., Barbazuk, B. & Douglas, C.J. (2007) Genome-wide analyses of phenylpropanoid-

- related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Canadian Journal of Botany*, **85**, 1182–1201.
- Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. & Legrand, M. (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry*, **278**, 95–103.
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M. et al. (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, **5**, 3978.
- Hu, Y., Gai, Y., Yin, L., Wang, X., Feng, C., Feng, L. et al. (2010) Crystal structures of a *Populus tomentosa* 4-coumarate:CoA ligase shed light on its enzymatic mechanisms. *The Plant Cell*, **22**, 3093–3104.
- Huerta-Cepas, J., Serra, F. & Bork, P. (2016) ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, **33**, 1635–1638.
- Jahns, P. & Holzwarth, A.R. (2012) The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1817**(1), 182–193.
- Jiao, C., Sørensen, I., Sun, X., Sun, H., Behar, H., Alseekh, S. et al. (2020) The *Penium margaritaceum* genome: hallmarks of the origins of land plants. *Cell*, **181**(P1097–1111), E12.
- Jones, P., Binns, D., Chang, H.-y., Fraser, M., Li, W., McAnulla, C. et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
- Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F. et al. (2015) Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nature Plants*, **1**, 14004.
- Jun, S.-Y., Sattler, S.A., Cortez, G.S., Vermerris, W., Sattler, S.E. & Kang, C. (2018) Biochemical and structural analysis of substrate specificity of a Phenylalanine Ammonia-Lyase. *Plant Physiology*, **176**, 1452–1468.
- Kalyaanamoorthy, S., Minh, B.Q., Wong, T., von Haeseler, A. & Jermini, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, **14**, 587–589.
- Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**, 772–780.
- Kaur, H., Heinzl, N., Schöttner, M., Baldwin, I.T. & Gális, I. (2010) R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. *Plant Physiology*, **152**, 1731–1747.
- Kim, R.J., Kim, H.J., Shim, D. & Suh, M.C. (2016) Molecular and biochemical characterizations of the monoacylglycerol lipase gene family of *Arabidopsis thaliana*. *The Plant Journal*, **85**, 758–771.
- Kim, S.-J., Kim, M.-R., Bedgar, D.L., Moinuddin, S.G.A., Cardenas, C.L., Davin, L.B. et al. (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 1455–1460.
- König, S., Feussner, K., Kaefer, A., Landesfeind, M., Thurow, C., Karlovsky, P. et al. (2014) Soluble phenylpropanoids are involved in the defense response of Arabidopsis against *Verticillium longisporum*. *New Phytologist*, **202**, 823–837.
- Kriegshauser, L., Knosp, S., Grienberger, E., Tatsumi, K., Gütle, D.D., Sørensen, I. et al. (2021) Function of the HYDROXYCINNAMOYL-CoA: SHIKIMATE HYDROXYCINNAMOYL TRANSFERASE is evolutionarily conserved in embryophytes. *The Plant Cell*, **33**, 1472–1491. <https://doi.org/10.1093/plcell/koab044>
- Labar, G., Bauvois, C., Borel, F., Ferrer, J.-L., Wouters, J. & Lambert, D.M. (2010) Crystal structure of the human Monoacylglycerol Lipase, a key actor in endocannabinoid signaling. *ChemBioChem*, **11**, 218–227.
- Labeeuw, L., Martone, P.T., Boucher, Y. & Case, R.J. (2015) Ancient origin of the biosynthesis of lignin precursors. *Biology Direct*, **10**, 23.
- Lacombe, E., Hawkins, S., Doorselaere, J.V., Piquemal, J., Goffner, D., Poeydomenge, O. et al. (1997) Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. *The Plant Journal*, **11**, 429–441.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R. et al. (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*, **40**, D1202–D1210.
- Lang, D., Ullrich, K.K., Murat, F. et al. (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, **93**, 515–533.
- Le, S.Q. & Gascuel, O. (2008) An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, **25**, 1307–1320.
- Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzenanner, M.A., Graham, S.W. et al. (2019) One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, **574**, 679–685.
- Letunic, I. & Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, **47**, W256–W259.
- Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M. et al. (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, **4**, 460–472.
- Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z. et al. (2020) *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants*, **6**, 259–272.
- Li, W., Zhang, F., Wu, R., Jia, R., Li, G., Guo, Y. et al. (2017) A novel N-Methyltransferase in Arabidopsis appears to feed a conserved pathway for nicotinate detoxification among land plants and is associated with lignin biosynthesis. *Plant Physiology*, **174**, 1492–1504.
- Louie, G.V., Bowman, M.E., Tu, Y., Mouradov, A., Spangenberg, G. & Noel, J.P. (2010) Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. *The Plant Cell*, **22**, 4114–4127.
- Maeda, H.A. & Fernie, A.R. (2021) Evolutionary history of plant metabolism. *Annual Review of Plant Biology*, **72**, 185–216. <https://doi.org/10.1146/annurev-arplant-080620-031054>
- Mansell, R.L.G., Gross, G.G., Stöckigt, J., Franke, H. & Zenk, M.H. (1974) Purification and properties of cinnamyl alcohol dehydrogenase from higher plants involved in lignin biosynthesis. *Phytochemistry*, **13**, 2427–2435.
- Martone, P.T., Estevez, J.M., Lu, F., Ruel, K., Denny, M.W., Somerville, C. et al. (2009) Discovery of lignin in seaweed reveals convergent evolution of cell-wall architecture. *Current Biology*, **19**, 169–175.
- Martz, F., Maury, S., Pincon, G. & Legrand, M. (1998) cDNA cloning, substrate specificity and expression study of tobacco caffeoyl-CoA 3-O-methyltransferase, a lignin biosynthetic enzyme. *Plant Molecular Biology*, **36**, 427–437.
- Matsuno, M., Compagnon, V., Schoch, G.A., Schmitt, M., Debayle, D., Bassard, J.-E. et al. (2009) Evolution of a novel phenolic pathway for pollen development. *Science*, **325**, 1688–1692.
- Maury, S., Geoffroy, P. & Legrand, M. (1999) Tobacco O-methyltransferases involved in phenylpropanoid metabolism: the different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. *Plant Physiology*, **121**, 215–224.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B. et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, **318**, 245–250.
- Miller, M., Owens, S.J. & Rørslett, B. (2011) Plants and colour: flowers and pollination. *Optics & Laser Technology*, **43**, 282–294.
- Moffitt, M.C., Louie, G.V., Bowman, M.E., Pence, J., Noel, J.P. & Moore, B.S. (2007) Discovery of two cyanobacterial phenylalanine ammonia lyases: kinetic and structural characterization. *Biochemistry*, **46**, 1004–1012.
- Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N. et al. (2012) Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology*, **13**, R74.
- Nagy, E.Z.A., Tork, S.D., Lang, P.A., Filip, A., Irimie, F.D., Poppe, L. et al. (2019) Mapping the hydrophobic substrate binding site of Phenylalanine Ammonia-Lyase from *Petroselinum crispum*. *ACS Catalysis*, **9**, 8825–8834.
- Nakatsu, T., Ichiyama, S., Hiratake, J., Saldanha, A., Kobashi, N., Sakata, K. et al. (2006) Structural basis for the spectral difference in luciferase bioluminescence. *Nature*, **440**, 372–376.
- Nelson, D. & Werck-Reichhart, D. (2011) A P450-centric view of plant evolution. *The Plant Journal*, **66**, 194–211.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular biology and evolution*, **32**, 268–274.

- Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K. *et al.* (2018) The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell*, **174**, 448–464.
- Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G. *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**, 579–584.
- Oliva, J., Rommel, S., Fossdal, C.G., Hietala, A.M., Nemesio-Gorrioz, M., Solheim, H. *et al.* (2015) Transcriptional responses of Norway spruce (*Picea abies*) inner sapwood against *Heterobasidion parviporum*. *Tree Physiology*, **35**, 1007–1015.
- Omura, T. (1999) Forty years of cytochrome P450. *Biochemical and Biophysical Research Communications*, **266**, 690–698.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K. *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research*, **35**, D883–D887.
- Overdijk, E.J.R., de Keizer, J., de Groot, D., Schoina, C., Bouwmeester, K., Ketelaar, T. *et al.* (2016) Interaction between the moss *Physcomitrella patens* and *Phytophthora*: a novel pathosystem for live-cell imaging of subcellular defence. *Journal of Microscopy*, **263**, 171–180.
- Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N. *et al.* (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 7705–7710.
- Pan, H., Zhou, R., Louie, G.V., Mühlemann, J.K., Bomati, E.K., Bowman, M.E. *et al.* (2014) Structural Studies of Cinnamoyl-CoA Reductase and Cinnamyl-Alcohol Dehydrogenase, key enzymes of monolignol biosynthesis. *The Plant Cell*, **26**, 3709–3727.
- Piatkowski, B.T., Imwattana, K., Tripp, E.A., Weston, D.J., Healey, A., Schmutz, J. *et al.* (2020) Phylogenomics reveals convergent evolution of red-violet coloration in land plants and the origins of the anthocyanin biosynthetic pathway. *Molecular Phylogenetics and Evolution*, **151**, 106904.
- Ponce De León, I., Schmelz, E.A., Gaggero, C., Castro, A., Álvarez, A. & Montesano, M. (2012) *Physcomitrella patens* activates reinforcement of the cell wall, programmed cell death and accumulation of evolutionary conserved defence signals, such as salicylic acid and 12-oxo-phytodienoic acid, but not jasmonic acid, upon *Botrytis cinerea* infection. *Molecular Plant Pathology*, **13**, 960–974.
- Prochnik, S.e., Umen, J., Nedelcu, A.m., Hallmann, A., Miller, S.m., Nishii, I. *et al.* (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carterii*. *Science*, **329**, 223–226.
- Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P. *et al.* (2018) The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology*, **28**, 733–745.
- Ralph, J., Lundquist, K., Brunow, G., Lu, F., Kim, H., Schatz, P.F. *et al.* (2004) Lignins: natural polymers from oxidative coupling of 4-hydroxyphenylpropanoids. *Phytochemistry Reviews*, **3**, 29–60.
- Renault, H., Alber, A., Horst, N.A., Basilio Lopes, A., Fich, E.A., Kriegshauser, L. *et al.* (2017a) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nature Communications*, **8**, 14713.
- Renault, H., De Marothy, M., Jonasson, G., Lara, P., Nelson, D.R., Nilsson, IngMarie *et al.* (2017b) Gene duplication leads to altered membrane topology of a cytochrome P450 enzyme in seed plants. *Molecular Biology and Evolution*, **34**, 2041–2056.
- Renault, H., Werck-Reichhart, D. & Weng, J.-K. (2019) Harnessing lignin evolution for biotechnological applications. *Current Opinion in Biotechnology*, **56**, 105–111.
- Rensing, S.A. (2014) Gene duplication as a driver of plant morphogenetic evolution. *Current Opinion in Plant Biology*, **17**, 43–48.
- Rensing, S.A. (2018) Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology*, **42**, 49–54.
- Rippin, M., Becker, B. & Holzinger, A. (2017) Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant and Cell Physiology*, **58**, 2067–2084.
- Rippin, M., Pichrtová, M., Arc, E., Kranner, I., Becker, B. & Holzinger, A. (2019) Metatranscriptomic and metabolite profiling reveals vertical heterogeneity within a *Zygnema* green algal mat from Svalbard (High Arctic). *Environmental Microbiology*, **21**, 4283–4299.
- Ro, D.K., Mah, N., Ellis, B.E. & Douglas, C.J. (2001) Functional characterization and subcellular localization of poplar (*Populus trichocarpa* × *Populus deltoides*) cinnamate 4-hydroxylase. *Plant Physiology*, **126**, 317–329.
- Rupasinghe, S., Baudry, J. & Schuler, M.A. (2003) Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. *Protein Engineering*, **16**, 721–731.
- Russell, D.W. & Conn, E.E. (1967) The cinnamic acid 4-hydroxylase from pea seedlings. *Archives of Biochemistry and Biophysics*, **122**, 256–258.
- Saballos, A., Ejeta, G., Sanchez, E., Kang, C. & Vermerris, W. (2009) A genome-wide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the brown midrib6 gene. *Genetics*, **181**, 783–795.
- Saito, N. & Harborne, J.B. (1992) Correlations between anthocyanin type, pollinator and flower colour in the labiatae. *Phytochemistry*, **31**, 3009–3015.
- Scheres, B. & van der Putten, W.H. (2017) The plant perceptron connects environment to development. *Nature*, **543**, 337–345.
- Schoenbohm, C., Martens, S., Eder, C., Forkmann, G. & Weisshaar, B. (2000) Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biological Chemistry*, **381**, 749–753.
- Schwede, T.F., Rétey, J. & Schulz, G.E. (1999) Crystal structure of Histidine Ammonia-Lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry*, **38**, 5355–5361.
- Sheahan, J.J. (1996) Sinapate esters provide greater UV-B attenuation than flavonoids in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany*, **83**, 679–686.
- Sheehan, H., Hermann, K. & Kuhlemeier, C. (2012) Color and scent: how single genes influence pollinator attraction. *Cold Spring Harbor Symposia on Quantitative Biology*, **77**, 117–133.
- Shockey, J.M., Fulda, M.S. & Browse, J. (2003) Arabidopsis contains a large superfamily of Acyl-Activating Enzymes. Phylogenetic and biochemical analysis reveals a new class of Acyl-Coenzyme A Synthetases. *Plant Physiology*, **132**, 1065–1076.
- Sierro, N., Battey, J., Ouadi, S., Bakaher, N., Bovet, L., Willig, A. *et al.* (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nature Communications*, **5**, 3833.
- Slotte, T., Hazzouri, K.M., Ågren, J.A., Koenig, D., Maumus, F., Guo, Y.-L. *et al.* (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nature Genetics*, **45**, 831–835.
- Sørensen, I., Pettolino, F.A., Bacic, A., Ralph, J., Lu, F., O'Neill, M.A. *et al.* (2011) The charophycean green algae provide insights into the early origins of plant cell walls. *The Plant Journal*, **68**, 201–211.
- Suzuki, S. & Umezawa, T. (2007) Biosynthesis of lignans and norlignans. *Journal of Wood Science*, **53**, 273–284.
- Sytar, O., Zivcak, M., Bruckova, K., Brestic, M., Hemmerich, I., Rauh, C. *et al.* (2018) Shift in accumulation of flavonoids and phenolic acids in lettuce attributable to changes in ultraviolet radiation and temperature. *Scientia Horticulturae*, **239**, 193–204.
- Szövényi, P., Frangedakis, E., Ricca, M., Quandt, D., Wicke, S. & Langdale, J.A. (2015) Establishment of *Anthoceros agrestis* as a model species for studying the biology of hornworts. *BMC Plant Biology*, **15**, 1–7.
- Szövényi, P., Gunadi, A. & Li, F.-W. (2021) Charting the genomic landscape of seed-free plants. *Nature Plants*, **7**, 554–565.
- Tan, D.-X., Hardeland, R., Manchester, L.C., Korkmaz, A., Ma, S., Rosales-Corral, S. *et al.* (2012) Functional roles of melatonin in plants, and perspectives in nutritional and agricultural science. *Journal of Experimental Botany*, **63**, 577–597.
- Tanaka, A., Shigemitsu, T., Yokota, Y. & Shika, N. (1997) A new *Arabidopsis* mutant induced synthesis with spotted pigmentation. *Genes & Genetic Systems*, **72**, 141–148.
- The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.
- Urban, P., Werck-Reichhart, D., Teutsch, H.G., Durst, F., Regnier, S., Kazmier, M. *et al.* (1994) Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast. Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway. *European Journal of Biochemistry*, **222**, 843–850.
- Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G. *et al.* (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science*, **341**, 1103–1106.

- Vanholme, R., De Meester, B., Ralph, J. & Boerjan, W. (2019) Lignin biosynthesis and its integration into metabolism. *Current Opinion in Biotechnology*, **56**, 230–239.
- Vanholme, R., Storme, V., Vanholme, B., Sundin, S., Christensen, J.H., Goemine, G. et al. (2012) A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *The Plant Cell*, **24**, 3506–3529.
- Vogt, T. (2010) Phenylpropanoid biosynthesis. *Molecular Plant*, **3**, 2–20.
- Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Geddes, B., Lu, F. et al. (2015) Syringyl lignin production in conifers: Proof of concept in a Pine tracheary element system. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6218–6223.
- Wan, T., Liu, Z.-M., Li, L.-F., Leitch, A.R., Leitch, I.J., Lohaus, R. et al. (2018) A genome for gnetophytes and early evolution of seed plants. *Nature Plants*, **4**, 82–89.
- Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y. et al. (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants*, **6**, 95–106.
- Weng, J.K. (2013) The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist*, **201**, 1141–1149.
- Weng, J.-K., Akiyama, T., Ralph, J. & Chapple, C. (2011) Independent recruitment of an O-methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*. *The Plant Cell*, **23**, 2708–2724.
- Weng, J.-K. & Chapple, C. (2010) The origin and evolution of lignin biosynthesis. *New Phytologist*, **187**, 273–285.
- Weng, J.-K., Li, X., Stout, J. & Chapple, C. (2008) Independent origins of syringyl lignin in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 7887–7892.
- Whelan, S. & Goldman, N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*, **18**, 691–699.
- Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N. et al. (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4859–E4868.
- Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M. et al. (2011) Origin of land plants: do conjugating green algae hold the key? *BMC Evolutionary Biology*, **11**, 104.
- Wohl, J. & Petersen, M. (2020) Functional expression and characterization of cinnamic acid 4-hydroxylase from the hornwort *Anthoceros agrestis* in *Physcomitrella patens*. *Plant Cell Reports*, **39**, 597–607.
- Wolf, L., Rizzini, L., Stracke, R., Ulm, R. & Rensing, S.A. (2010) The molecular and physiological responses of *Physcomitrella patens* to ultraviolet-B radiation. *Plant Physiology*, **153**, 1123–1134.
- Worden, A.z., Lee, J.-h., Mock, T., Rouze, P., Simmons, M.p., Aerts, A.I. et al. (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, **324**, 268–272.
- Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K.L. et al. (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics*, **10**, S3.
- Xue, J.-S., Zhang, B., Zhan, H., Lv, Y.-L., Jia, X.-L., Wang, T. et al. (2020) Phenylpropanoid derivatives are essential components of sporopollenin in vascular plants. *Molecular Plant*, **13**, 1644–1653.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. & Zhang, Y. (2015) The I-TASSER suite: protein structure and function prediction. *Nature Methods*, **12**, 7–8.
- Ye, Z.-H., Kneusel, R.E., Matern, U. & Varner, J.E. (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *The Plant Cell*, **6**, 1427–1439.
- Ye, Z.-H. & Varner, J.E. (1995) Differential expression of two O-methyltransferases in lignin biosynthesis in *Zinnia elegans*. *Plant Physiology*, **108**, 459–467.
- Youn, B., Camacho, R., Moinuddin, S.G.A., Lee, C., Davin, L.B., Lewis, N.G. et al. (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Organic & Biomolecular Chemistry*, **4**, 1687–1697.
- Zhang, J., Fu, X.-X., Li, R.-Q., Zhao, X., Liu, Y., Li, M.-H. et al. (2020) The hornwort genome and early land plant evolution. *Nature Plants*, **6**, 107–118.
- Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.
- Zubieta, C., Kota, P., Ferrer, J.L., Dixon, R.A. & Noel, J.P. (2002) Structural basis for the modulation of lignin monomer methylation by caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase. *The Plant Cell*, **14**, 1265–1277.

4.3 **Publication VI:** Different patterns of gene evolution underpin water-related innovations in land plants

This article is a Commentary on:

Bowles, AMC, Paps, J and Bechtold, U. (2022) Water-related innovations in land plants evolved by different patterns of gene cooption and novelty. *New Phytologist*, 235: 732-742. <https://doi.org/10.1111/nph.17981>

This article published in the Journal “New Phytologist” in July 2022. The full article can be found online: <https://doi.org/10.1111/nph.18176>

Contribution of Janine Fürst-Jansen, first author

J. M. R. Fürst-Jansen planned the manuscript and designed and prepared figure 1. She wrote parts of the manuscript and critically read and revised the manuscript.

Commentary

Different patterns of gene evolution underpin water-related innovations in land plants

About 500 million years ago (Ma) land plants emerged (Morris *et al.*, 2018). This was one of the most transformative events in Earth's history, giving rise to an astonishing diversity of species that formed complex plant-dominated ecosystems. The earliest land plants, therefore, needed to respond to the myriad of terrestrial stressors posed by the new environment. Perhaps the most obvious ones are those related to water availability, that is reduced access to water and increased water loss by drought and temperature stress. Stomata are a key trait of land plants; roots and vascular tissue further facilitated these adaptations, underpinning the success of Tracheophyta. These systems allow the active and more efficient absorption and transport of water and the regulation of its evaporation. In this issue of *New Phytologist*, Bowles *et al.* (2022; pp. 732–742) use available genomic and transcriptomic data across all major lineages of streptophytes to paint a picture of the evolutionary history that shaped the molecular toolkit required to develop and operate roots, vascular tissue and stomata.

'...gene duplication and co-option are more prevalent than gene novelty for the three morphological innovations under study.'

How the first land plants coped with the differences in water and nutrient availability and provided protection from evaporation is a central question in plant terrestrialisation. Stomata-controlled transpiration is a prime trait involved in water distribution and regulation. An increased water conductance was achieved with the appearance of the vasculature. Rhizoids are thought to have provided water uptake and anchorage to soil in early land plants. Roots of vascular plants with meristem-formed tissue can penetrate into greater depth of substrate, tapping new resources and forming large underground root architectures. The origin of roots was possible due to the emergence of complex vasculature in the common ancestor of vascular plants; this was corroborated by the fact that roots evolved twice in vascular plants (Hetherington & Dolan, 2018). Phloem and xylem are key in transporting water and nutrients. With the appearance of vasculature also came the ability

to grow erect, allowing for shade avoidance, the decoupling of environmental shape and accessibility to light. The series in which these morphological traits appeared during the evolution of land plants (Fig. 1) hints that not all of them appeared in the earliest land plants, but they certainly allowed the embryophyte diversity past and present. The recent increase in genomic and transcriptomic data, now covering all major representative lineages of streptophytes, allows us to investigate the evolutionary series by which molecular networks that underpin these morphological features have arisen (Harris *et al.*, 2020).

Comparing the genomes of land plants and streptophyte algae allows the inference of the genetic toolkits that might have been present in the ancestors that first colonised the terrestrial environment. Several previous studies have exploited this comparative genomics approach, finding two events of major gene emergences in the ancestors of streptophytes and embryophytes (Leebens-Mack *et al.*, 2019; Bowles *et al.*, 2020). Here, Bowles *et al.* (2022) build upon their earlier 2020 analyses and investigate the origin and evolution of roots, vascular tissue and stomata. They combined sequence-similarity search and phylogenomics to reconstruct the likely origin of gene families associated with root, vasculature and stomata development by assigning it to the most recent common ancestor of the species containing that homologue. If the inferred genetic ancestor corresponds to that where the trait of interest originated, it was considered to have emerged *de novo*. If genes predated the origin of the trait, the genes were considered to have been co-opted. If gene duplications were inferred at the node where the trait appeared, then gene duplication (probably followed by sub- and neofunctionalisation) was invoked as responsible for the gene–phenotype association. Indeed, Bowles *et al.* (2022) found that gene duplication and co-option were more prevalent than *de novo* gene emergence for the three morphological innovations under study. This amalgamates well with both many previous large-scale *in silico* analyses – such as phylostrata analyses of gene expression networks (Ruprecht *et al.*, 2017) that showed certain traits to be accompanied by duplication and network-rewiring – as well as a diversity of wet-laboratory data, including the evolution of ROOTHAIRLESS LIKE transcription factors that are required for the development of rhizoids in *Marchantia polymorpha* and of root hairs in *Arabidopsis thaliana* (Breuninger *et al.*, 2016).

The comparative approach of genomic data is a powerful one; it allows the inference of historical (evolutionary) events through the comparison of present-day genomes within a robust phylogenetic framework. Two major factors influence the robustness of comparative genomics: (1) the representation of the phylogenetic diversity of a group under consideration; and (2) the robustness of the phylogenetic framework onto which the results are mapped. Thanks to the recent availability of plant genomes (and transcriptomes) we can tackle key questions in their evolution through comparative genomics. Ongoing genome sequencing efforts are likely to

This article is a Commentary on Bowles *et al.* (2022), 235: 732–742.

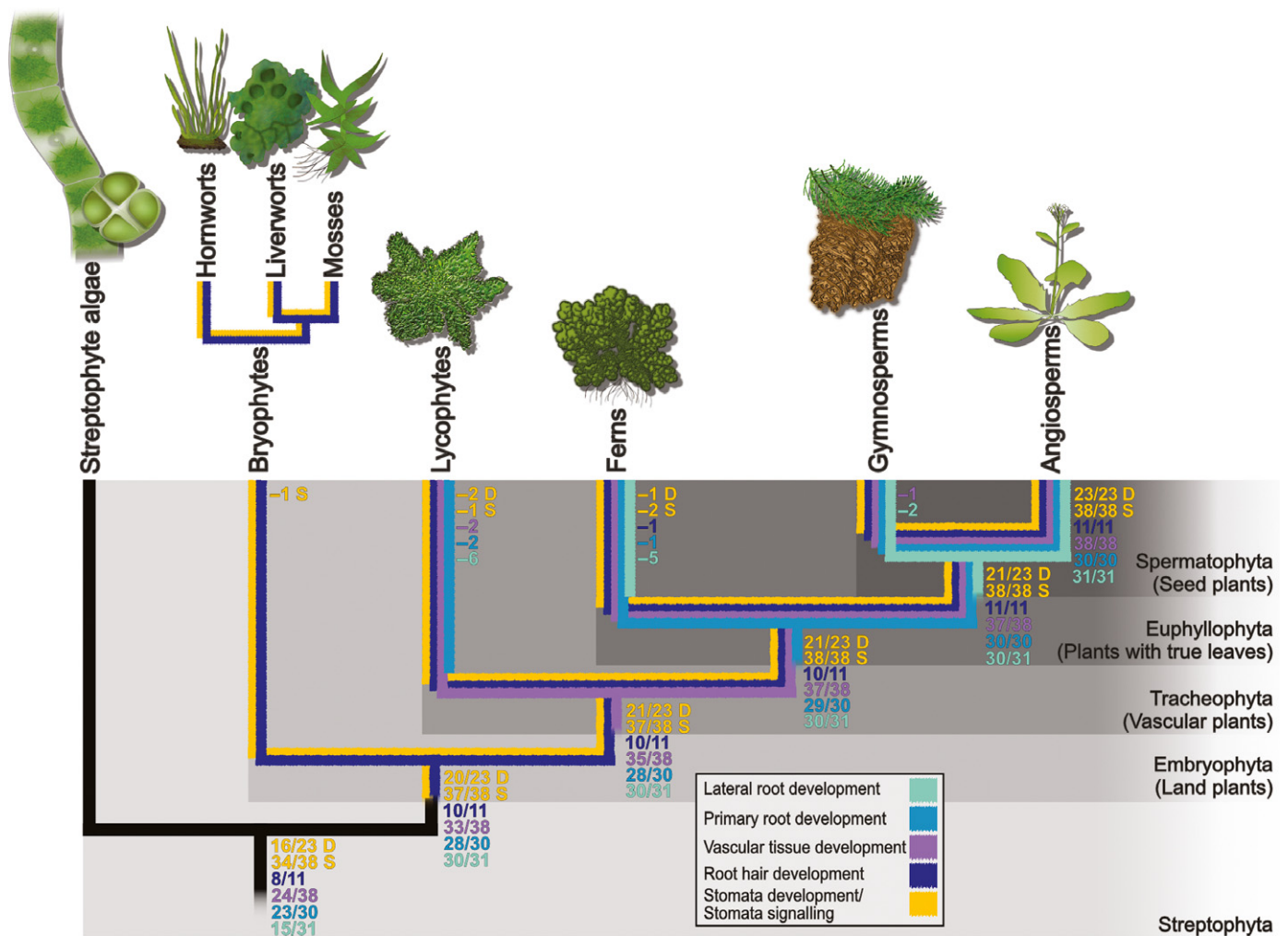


Fig. 1 A simplified overview of the evolution of water-related innovations in streptophytes. Coloured lines indicate the evolution of stomata development and signalling (yellow), root hairs (dark blue), primary (teal) and lateral roots (light blue), and vascular tissue (light purple). Numbers at internal nodes represent the inferred presence of homology groups (HG) with respect to the total number of investigated HGs for the respective traits. Numbers are derived from the presence/absence heat maps shown in this issue of *New Phytologist* by Bowles *et al.* (2022; pp. 732–742) and do not represent lineage-specific gains and losses, but are cumulative numbers of how many HGs for a certain trait are predicted to be present at the last common ancestor of a monophyletic group. HGs that remained undetected in specific lineages are denoted with (–). For HGs associated with stomata, we distinguish between development (D) and signalling (S). Illustrations represent the major lineages of streptophytes: *Chlorokybus cerfii* and *Zygnema circumcarinatum* (streptophyte algae); *Anthoceros agrestis* (hornworts); *Marchantia polymorpha* (liverworts); *Physcomitrium patens* (mosses); *Selaginella moellendorffii* (lycophytes); *Azolla filiculoides* (ferns); *Picea abies* (gymnosperms); and *Arabidopsis thaliana* (angiosperms).

empower future comparative genomic studies with a phylogenetically more diverse set of chromosome-resolved genomes.

Bowles *et al.* (2022) take advantage of the large transcriptomic dataset derived from the 1000 plants (1KP) project (Leebens-Mack *et al.*, 2019) to improve the representation of the phylogenetic diversity of plants. While transcriptomes cannot represent the full genomic make-up of a species, homologues derived from transcriptomic data are unmistakable evidence of their presence, thereby increasing the resolution of the identified patterns of gene origin and duplication. A robust phylogenetic framework allows us to map traits and discern convergence from a common descent. The burst of phylogenomic studies over the last few years has completely changed our view of plant evolution. The interrelationships of streptophyte algal lineages were updated, identifying the Zygnematophyceae as the closest living relatives of land plants and supporting the

monophylum of bryophytes (hornworts, liverworts, mosses) as sister to all vascular plants (Puttick *et al.*, 2018; Leebens-Mack *et al.*, 2019). This has important implications for the evolution of several key traits, including the origin of stomata studied here, which are likely to have originated in the last common ancestor of land plants and was secondarily lost in some liverworts and mosses (Harris *et al.*, 2020). Among the genes tied to stomata function, Bowles *et al.* (2022) investigated abscisic acid (ABA) signalling, including the three subfamilies of the ABA receptor PYRABACTIN RESISTANCE 1-LIKE (PYR1/PYL) and gene families for the components of its downstream signalling cascade, 2C-type protein phosphatases (PP2C) and sucrose nonfermenting 1-related kinase 2 (SnRK2.1-10) (Cutler *et al.*, 2010). Homology groups for PYL, PP2C, and SnRK were identified in the streptophyte ancestor (Bowles *et al.*, 2022). This is congruent with earlier analyses: PYL

orthologues have been reported for streptophyte algae (de Vries *et al.*, 2018; Cheng *et al.*, 2019). A functional characterisation by Sun *et al.* (2019) has highlighted that, while a zygnetomophyceae PYL homologue is able to inhibit the downstream PP2C phosphatase, this inhibition occurs in an ABA-independent manner. Additionally, the downstream cascade of SnRK2 signalling appears to have emerged even earlier in streptophyte evolution (Lind *et al.*, 2015). This showcases the fact that the genetic make-up of ABA-dependent stomatal regulation has deep evolutionary roots, but it was integrated to stomatal control only later in land plant evolution. A similar picture emerges for the genetic regulation of vascular tissue development: Xu *et al.* (2014) hypothesised deep homology between the xylem and phloem of vascular plants and the water- and nutrient-conducting cells of bryophytes. Hydroid differentiation appears to be controlled by the same group of NAC transcription factors that also regulate xylem patterning.

In targeted comparative genomics studies, in which genes that are associated with a certain trait are specifically studied, the strength of functional annotation is central. When talking about plants, most genes have been functionally characterised in the model species *Arabidopsis thaliana* and functional conservation is predicted based on homology and the conservation of key functional residues (de Vries *et al.*, 2021). The benefits and limitations of an *Arabidopsis*-centric functional characterisation of plant genes is nicely discussed by Bowles *et al.* (2022). Evolutionary inferences on the genetic toolkit underlying key traits of interest would greatly benefit from functional studies in algae and other nonseed and nonvascular plants. For example, in the synthesis of lignin, which is part of the backbone of the vasculature, the lycophyte *Selaginella moellendorffii* and the angiosperm *Arabidopsis thaliana* have convergently evolved the functionalities of the enzymes ferulate 5-hydroxylase (F5H) and caffeic acid *O*-methyltransferase (COMT) (Weng *et al.*, 2008, 2011). Ideally, large-scale comparative studies should move from the study of the evolutionary history of certain gene families to a more realistic view of larger evolutionary patterns that accompanied the acquisition of, for example, evolutionary developmental biology (evo-devo) traits.


Taken together, comparative studies of genomic data allow us to identify new general evolutionary patterns. Specifically how broadly distributed are these patterns – not only in terms of gene families but also regarding diversity – is impossible to calculate with smaller scale genetic studies. However, large-scale comparative genomics build upon the functional analyses of genes and therefore need primary data acquired through slow and painstaking work of, for example, taxonomy or functional genetic studies for which there is no shortcut. More so, genetic characterisations focus on the few genetically accessible organisms making our picture of diversity in genetic solutions limited. Ideally, a diverse set of model species should be combined. In this study, the most prevalent evolutionary pattern is that the genetic background underlying many of the land plant innovations that has much older evolutionary roots than the traits themselves. Gene duplication accompanied by neo- or subfunctionalisation have been highlighted and co-option has been hypothesised as major patterns by which the genetic networks underpinning these traits have evolved. This seems a recurrent pattern for the origin of several innovations in land plants

(Breuninger *et al.*, 2016; Ruprecht *et al.*, 2017; Bowles *et al.*, 2020), supporting the idea that genetic rewiring of developmental and stress response pathways might be more prevalent than the origin of new genes for generating complexity, as proposed for the evolution of stress signalling networks during streptophyte evolution (Fürst-Jansen *et al.*, 2020; Rieseberg *et al.*, 2022).


Acknowledgements

We thank Prof. Dr Jan de Vries for proofreading. JMRF-J is supported by the 'Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular Biosciences' (GGNB), University of Göttingen. JMRF-J and II acknowledge financial support from the European Research Council (Grant no. 852725; ERC-StG 'TerreStriAL' to Jan de Vries, University of Goettingen). JMRF-J and II are part of the MADLand DFG priority programme 2237 (<http://madland.science>).

ORCID

Janine M. R. Fürst-Jansen  <https://orcid.org/0000-0002-5269-8725>

Iker Irisarri  <https://orcid.org/0000-0002-3628-1137>

Sophie Vries de  <https://orcid.org/0000-0002-5267-8935>

Janine M. R. Fürst-Jansen^{1*} , Sophie de Vries¹  and Iker Irisarri^{1,2} 

¹Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Goettingen, Goldschmidtstr. 1, 37077 Goettingen, Germany;

²Campus Institute Data Science (CIDAS), University of Goettingen, Goldschmidtstr. 1, 37077 Goettingen, Germany

(*Author for correspondence: email janine.fuerst-jansen@uni-goettingen.de)

References

- Bowles AMC, Bechtold U, Paps J. 2020. The origin of land plants is rooted in two bursts of genomic novelty. *Current Biology* 30: 530–536.e2.
- Bowles AMC, Paps J, Bechtold U. 2022. Water-related innovations in land plants evolved by different patterns of gene cooption and novelty. *New Phytologist* 235: 732–742.
- Breuninger H, Thamm A, Streubel S, Sakayama H, Nishiyama T, Dolan L. 2016. Diversification of a transcription factor family led to the evolution of antagonistically acting genetic regulators of root hair growth. *Current Biology* 26: 1622–1628.
- Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang YU *et al.* 2019. Genomes of subaerial Zygnematomphyceae provide insights into land plant evolution. *Cell* 179: 1057–1067.
- Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. 2010. Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology* 61: 651–679.
- Fürst-Jansen JMR, de Vries S, de Vries J. 2020. Evo-physio: on stress responses and the earliest land plants. *Journal of Experimental Botany* 71: 3254–3269.
- Harris BJ, Harrison CJ, Hetherington AM, Williams TA. 2020. Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Current Biology* 30: 2001–2012.
- Hetherington AJ, Dolan L. 2018. Stepwise and independent origins of roots among land plants. *Nature* 561: 235–238.

- Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S *et al.* 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574: 679–685.
- Lind C, Dreyer I, López-Sanjurjo E, von Meyer K, Ishizaki K, Kohchi T, Lang D, Zhao Y, Kreuzer I, Al-Rasheid K *et al.* 2015. Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure. *Current Biology* 25: 928–935.
- Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* 115: E2274–E2283.
- Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, Pressel S, Wellman CH, Schneider H, Pisani D *et al.* 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology* 28: 733–745.
- Rieseberg TP, Dadras A, Fürst-Jansen JMR, Dhabalia Ashok A, Darienko T, de Vries S, Irisarri I, de Vries J. 2022. Crossroads in the evolution of plant specialized metabolism. *Seminars in Cell & Developmental Biology*. doi: 10.1016/j.semcdb.2022.03.004.
- Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, Becker JD, Vandepoele K, Mutwil M. 2017. Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *The Plant Journal* 90: 447–465.
- Sun Y, Harpazi B, Wijerathna-Yapa A, Merilo E, de Vries J, Michaeli D, Gal M, Cuming AC, Kollist H, Mosquna A. 2019. A ligand-independent origin of abscisic acid perception. *Proceedings of the National Academy of Sciences, USA* 116: 24892–24899.
- de Vries J, Curtis BA, Gould SB, Archibald JM. 2018. Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proceedings of the National Academy of Sciences, USA* 115: E3471–E3480.
- de Vries S, Fürst-Jansen JMR, Irisarri I, Ashok AD, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J. 2021. The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *The Plant Journal* 107: 975–1002.
- Weng J-K, Akiyama T, Ralph J, Chapple C. 2011. Independent recruitment of an O-methyltransferase for syringyl lignin biosynthesis in *Selaginella moellendorffii*. *Plant Cell* 23: 2708–2724.
- Weng J-K, Li X, Stout J, Chapple C. 2008. Independent origins of syringyl lignin in vascular plants. *Proceedings of the National Academy of Sciences, USA* 105: 7887–7892.
- Xu BO, Ohtani M, Yamaguchi M, Toyooka K, Wakazaki M, Sato M, Kubo M, Nakano Y, Sano R, Hiwatashi Y *et al.* 2014. Contribution of NAC transcription factors to plant adaptation to land. *Science* 343: 1505–1508.

Key words: comparative genomics, Embryophyta, evolution, phloem, root, stomata, xylem.

4.4 Pre-print VII: Chromosome-level genomes of multicellular algal sisters to land plants illuminate signaling network evolution

This manuscript was uploaded to the pre-print server Biorxiv. The newest online version as well as all supplementary figures and supplementary datasets can be found under:

<https://doi.org/10.1101/2023.01.31.526407>

Contribution of Janine Fürst-Jansen, middle author

J. M. R. Fürst-Jansen planned and performed dozens of stress experiments on *Zygnema circumcarinatum* SAG 698-1b (e.g., shown in figure 3, the supplement, and being the foundation for the wealth of data needed for the co-expression network analysis). This involved establishing and troubleshooting the RNA extraction protocol for *Zygnema*. She performed RNA isolation on the stress-treated samples and performed a quality check before sending the samples to the sequencing facility. She wrote parts of the manuscript and critically read and revised the manuscript.

Chromosome-level genomes of multicellular algal sisters to land plants illuminate signaling network evolution

Authors: Xuehuan Feng^{*1}, Jinfang Zheng^{*1}, Iker Irisarri^{*2,3,4}, Huihui Yu¹⁴, Bo Zheng¹, Zahin Ali³⁰, Sophie de Vries², Jean Keller⁵, Janine M.R. Fürst-Jansen², Armin Dadras², Jaccoline M.S. Zegers², Tim P. Rieseberg², Amra Dhabalia Ashok², Tatyana Darienko², Maaike J. Bierenbroodspot², Lydia Gramzow⁶, Romy Petroll^{7,8}, Fabian B. Haas^{7,8}, Noe Fernandez-Pozo^{7,9}, Orestis Nousias¹, Tang Li¹, Elisabeth Fitzek¹⁰, W. Scott Grayburn¹¹, Nina Rittmeier¹², Charlotte Permann¹², Florian Rümpler⁶, John M. Archibald¹³, Günter Theißen⁶, Jeffrey P. Mower¹⁴, Maike Lorenz¹⁵, Henrik Buschmann¹⁶, Klaus von Schwartzberg¹⁷, Lori Boston¹⁸, Richard D. Hayes¹⁹, Chris Daum¹⁹, Kerrie Barry¹⁹, Igor V. Grigoriev^{19,20,21}, Xiyin Wang²², Fay-Wei Li^{23,24}, Stefan A. Rensing^{7,25}, Julius Ben Ari²⁶, Noa Keren²⁶, Assaf Mosquna²⁶, Andreas Holzinger¹², Pierre-Marc Delaux⁵, Chi Zhang^{14,27}, Jinling Huang^{28,29}, Marek Mutwil³⁰, Jan de Vries^{#§2,3,31}, Yanbin Yin^{#§1}

1—University of Nebraska-Lincoln, Department of Food Science and Technology, Lincoln, NE 68588, USA

2—University of Goettingen, Institute of Microbiology and Genetics, Department of Applied Bioinformatics, Goldschmidtstr. 1, 37077 Goettingen, Germany

3—University of Goettingen, Campus Institute Data Science (CIDAS), Goldschmidtstr. 1, 37077 Goettingen, Germany

4—Section Phylogenomics, Centre for Molecular biodiversity Research, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Zoological Museum Hamburg, Martin-Luther-King-Platz 3, 20146 Hamburg, Germany

5—Laboratoire de Recherche en Sciences Végétales (LRSV), Université de Toulouse, CNRS, UPS, INP Toulouse, Castanet-Tolosan, 31326, France

6—University of Jena, Matthias Schleiden Institute / Genetics, 07743, Jena, Germany

7—Plant Cell Biology, Department of Biology, University of Marburg, Marburg, Germany

8—Department of Algal Development and Evolution, Max Planck Institute for Biology Tübingen, Tübingen, Germany

9—Institute for Mediterranean and Subtropical Horticulture “La Mayora” (UMA-CSIC)

10—Computational Biology, Department of Biology, Center for Biotechnology, Bielefeld University, Bielefeld, Germany

11—Northern Illinois University, Molecular Core Lab, Department of Biological Sciences, DeKalb, IL 60115, USA

12—University of Innsbruck, Department of Botany, Research Group Plant Cell Biology, Sternwartestraße 15, A-6020 Innsbruck, Austria

13—Dalhousie University, Department of Biochemistry and Molecular Biology, 5850 College Street, Halifax NS B3H 4R2, Canada

14—University of Nebraska-Lincoln, Center for Plant Science Innovation, Lincoln, NE 68588, USA

15—University of Goettingen, Albrecht-von-Haller-Institute for Plant Sciences, Experimental Phycology and Culture Collection of Algae at Goettingen University (EPSAG), Nikolausberger Weg 18, 37073 Goettingen, Germany

16—University of Applied Sciences Mittweida, Faculty of Applied Computer Sciences and Biosciences, Section Biotechnology and Chemistry, Molecular Biotechnology, Technikumplatz 17, 09648 Mittweida, Germany

17—Universität Hamburg, Institute of Plant Science and Microbiology, Microalgae and Zygnematophyceae Collection Hamburg (MZCH) and Aquatic Ecophysiology and Phycology, Ohnhorststr. 18, 22609, Hamburg, Germany

18—Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

19—Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

20—Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

21—Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA 94720, USA

22—North China University of Science and Technology

23—Boyce Thompson Institute, Ithaca, NY, USA

24—Cornell University, Plant Biology Section, Ithaca, NY, USA

25—University of Freiburg, Centre for Biological Signalling Studies (BIOS), Freiburg, Germany

26—The Hebrew University of Jerusalem, The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Rehovot 7610000, Israel

27—University of Nebraska-Lincoln, School of Biological Sciences, Lincoln, NE 68588, USA

28—State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China

29—Department of Biology, East Carolina University, Greenville, NC, USA

30—Nanyang Technological University, School of Biological Sciences, 60 Nanyang Drive, Singapore 637551, Singapore

31—University of Goettingen, Goettingen Center for Molecular Biosciences (GZMB), Justus-von-Liebig-Weg 11, 37077 Goettingen, Germany

*These authors contributed equally

#These authors contributed equally

§Authors for correspondence:

Jan de Vries, devries.jan@uni-goettingen.de

Yanbin Yin, yvin@unl.edu

57 **ORCID:**
58 Xuehuan Feng 0000-0002-0732-9322
59 Iker Irisarri 0000-0002-3628-1137
60 Huihui Yu 0000-0003-2725-1937
61 Sophie de Vries 0000-0002-5267-8935
62 Jean Keller 0000-0002-5198-0331
63 Janine MR Fürst-Jansen 0000-0002-5269-8725
64 Armin Dadras 0000-0001-7649-2388
65 Jacqueline Zegers 0000-0003-0042-8923
66 Tim P Rieseberg 0000-0003-3548-8475
67 Amra Dhabalia Ashok 0000-0001-5787-6941
68 Tatyana Darienko 0000-0002-1957-0076
69 Maaike J. Bierenbroodspot 0000-0002-4460-3394
70 Lydia Gramzow, 0000-0003-1919-6077
71 Romy Petroll, 0000-0003-0165-982X
72 Fabian B. Haas 0000-0002-7711-5282
73 Noe Fernandez-Pozo 0000-0002-6489-5566
74 Orestis Nousias, 0000-0001-6644-6599
75 Elisabeth Fitzek, 0000-0001-9423-2993
76 Charlotte Permann 0000-0003-0276-6053
77 John M. Archibald 0000-0001-7255-780X
78 Florian Rümpler 0000-0003-4986-2595
79 Günter Theißen: 0000-0003-4854-8692
80 Jeffrey P. Mower 0000-0002-1501-5809
81 Maike Lorenz 0000-0002-2277-3077
82 Henrik Buschmann 0000-0003-3022-6150
83 Klaus von Schwartzberg, 0000-0002-2273-9960
84 Richard Hayes 0000-0002-5236-7918
85 Chris Daum 0000-0003-3895-5892
86 Kerrie Barry 0000-0002-8999-6785
87 Igor Grigoriev 0000-0002-3136-8903
88 Fay-Wei Li 0000-0002-0076-0152
89 Stefan A. Rensing, 0000-0002-0225-873X
90 Assaf Mosquna, 0000-0003-0467-9287
91 Andreas Holzinger 0000-0002-7745-3978
92 Pierre-Marc Delaux 0000-0002-6211-157X
93 Chi Zhang 0000-0002-1827-8137
94 Jinling Huang, 0000-0003-4893-4171
95 Marek Mutwil 0000-0002-7848-0126
96 Jan de Vries, 0000-0003-3507-5195
97 Yanbin Yin, 0000-0001-7667-881X
98

99 **HIGHLIGHTS:**

- 100 1. Genomes of four filamentous algae (*Zygnema*) sisters to land plants
101 2. *Zygnema* are rich in genes for multicellular growth and environmental acclimation: signaling, lipid
102 modification, and transport
103 3. Cell wall innovations: diversification of hexameric rosette cellulose synthase in Zygnematophyceae
104 4. Co-expression networks reveal conserved modules for balancing growth and acclimation
105

106 **ABSTRACT**

107 The filamentous and unicellular algae of the class Zygnematophyceae are the closest algal relatives of
108 land plants. Inferring the properties of the last common ancestor shared by these algae and land plants
109 allows us to identify decisive traits that enabled the conquest of land by plants. We sequenced four
110 genomes of filamentous Zygnematophyceae (three strains of *Zygnema circumcarinatum* and one strain of
111 *Z. cylindricum*) and generated chromosome-scale assemblies for all strains of the emerging model system
112 *Z. circumcarinatum*. Comparative genomic analyses reveal expanded genes for signaling cascades,
113 environmental response, and intracellular trafficking that we associate with multicellularity. Gene family
114 analyses suggest that Zygnematophyceae share all the major enzymes with land plants for cell wall
115 polysaccharide synthesis, degradation, and modifications; most of the enzymes for cell wall innovations,
116 especially for polysaccharide backbone synthesis, were gained more than 700 million years ago. In
117 Zygnematophyceae, these enzyme families expanded, forming co-expressed modules. Transcriptomic
118 profiling of over 19 growth conditions combined with co-expression network analyses uncover cohorts of
119 genes that unite environmental signaling with multicellular developmental programs. Our data shed light
120 on a molecular chassis that balances environmental response and growth modulation across more than
121 600 million years of streptophyte evolution.
122

123 **INTRODUCTION**

124 Plant terrestrialization changed the surface of the Earth. In a fateful event about 550 million years ago, the
125 first representatives of the clade Embryophyta (land plants) gained a foothold on land (Delwiche and
126 Cooper, 2015; Morris et al., 2018; Fürst-Jansen et al., 2020; Strother and Foster, 2021; Harris et al.,
127 2022). These first land plants emerged from within the clade of Streptophyta that—until Embryophyta
128 emerged—consisted solely of streptophyte algae. Which lineage among these freshwater and terrestrial
129 algae is closest to land plants? Over the last decade, considerable advances have been made in
130 establishing a robust phylogenomic framework for streptophyte evolution and the birth of embryophytes
131 (Wodniok et al., 2011; Wickett et al., 2014; Puttick et al., 2018; Leebens-Mack et al., 2019). Moreover, a
132 long tradition of cell biological research in streptophyte algae is available (reviewed by Domozych and
133 Bagdan 2022). Both lines of evidence have come to the same conclusion: Zygnematophyceae are the
134 closest algal relatives of land plants.

135 Zygnematophyceae is a class of freshwater and semi-terrestrial algae with more than 4,000 described
136 species (Guiry 2021). Their unifying feature is sexual reproduction by conjugation and lack of motile
137 stages; zygospore formation can often be observed in field samples e.g., in *Mougeotia* (Permann et al.,
138 2021a) and *Spirogyra* (Permann et al., 2021b, 2022a). Zygnematophyceae have been recently rearranged
139 into five orders: Spirogloales, Serritaeniales, Spirogyrales, Desmidiiales, and Zygnematales (Hess et al.,
140 2022). Among these, filamentous growth likely evolved at least five times independently (Hess et al.,
141 2022). So far, genome sequences are only available for unicellular Zygnematophyceae: *Penium*
142 *margaritaceum* and *Closterium peracerosum–strigosum–littorale* in Desmidiiales (Jiao et al. 2020;

143 Sekimoto et al., 2023), *Mesotaenium endlicherianum* in Serritaeniales, and *Spirogloea muscicola* in
144 Spirogloales (both Cheng et al., 2019).

145 Zygnematophyceae possess adaptations to withstand terrestrial stressors. Some are protected against
146 desiccation by extracellular polymers like arabinogalactan proteins (AGPs; Palacio-Lopez et al. 2019) and
147 homogalacturonan-rich mucilage sheaths (Herburger et al., 2019). UV-absorbing phenolic compounds
148 were shown to occur in *Zygnema* (Pichrtová et al., 2013, Holzinger et al., 2018), *Zygogonium erictorum*
149 (e.g., Aigner et al., 2013, Herburger et al., 2016) and *Serritaenia* (Busch & Hess 2021). Recurrent
150 transcriptomic and metabolomic changes have been observed following desiccation (Rippin et al. 2017)
151 and other abiotic stresses (de Vries et al., 2018, Arc et al., 2020; Fitzek et al., 2019) in *Zygnema*, as well
152 as under natural conditions in the uppermost layers of Arctic *Zygnema* mats (Rippin et al. 2019). The
153 nature of these stress responses is of deep biological significance: various orthologous groups of proteins
154 once considered specific to land plants have recently been inferred to predate the origin of Embryophyta
155 (Nishiyama et al., 2018; Bowles et al., 2020). These include intricate transcription factor (TF) networks
156 (Wilhelmsson et al., 2017), phytohormone signaling pathways (Bowman et al., 2019), specialized
157 biochemical pathways (Rieseberg et al., 2022), symbiosis signaling (Delaux et al., 2015), and cell wall
158 modifications (Harholt et al., 2016).

159 Accurately inferring the developmental and physiological programs of the first land plant ancestor
160 depends on our ability to predict them in its sister group, the common ancestor of Zygnematophyceae.
161 Robustness of this inference lays on accounting for the phylogenetic, genomic, and morphologic diversity
162 of the Zygnematophyceae. Here, we report on the first four genomes of filamentous Zygnematophyceae
163 (Order Zygnematales), three from strains of *Zygnema circumcarinatum* and one from *Zygnema* cf.
164 *cylindricum*, which include the first chromosome-scale genomes for any streptophyte algae. Comparative
165 genomics using the new *Zygnema* genomes allow inferring the genetic repertoire of land plants ancestors
166 that first conquered the terrestrial environment. But which functional cohorts were relevant? Co-
167 expression analyses shed light on the deep evolutionary roots of the mechanism for balancing
168 environmental responses and multicellular growth.

169

170 **RESULTS AND DISCUSSION**

171

172 **First chromosome-level genomes for streptophyte algae**

173 The nuclear and organellar genomes of four *Zygnema* strains (*Z. circumcarinatum* SAG 698-1b, UTEX
174 1559, and UTEX 1560 and *Zygnema* cf. *cylindricum* SAG 698-1a, **Figure 1A-C**) were assembled using a
175 combination of PacBio High-Fidelity (HiFi) long reads, Oxford Nanopore long reads, and Illumina short
176 reads. In total, we sequenced 51 gigabases (Gb) (797X), 69 Gb (1042X), 6.7 Gb (103X), and 253 Gb
177 (786X) for strains SAG 698-1b, UTEX 1559, UTEX 1560, and SAG 698-1a, respectively (**Table S1A,B**).
178 Using chromatin conformation data (Dovetail Hi-C), we scaffolded the *Z. circumcarinatum* SAG 698-1b
179 genome (N50=4Mb; **Table 1**) into 90 final scaffolds; 98.6% of the assembly belongs to the 20 longest
180 scaffolds (**Table S1C**) corresponding to 20 pseudo-chromosomes (**Figure 1D**). Cytological chromosome
181 counting at different stages of mitosis (prophase, metaphase, telophase) (**Figure 1B, Figure S1**) verified
182 the 20 chromosomes, similar to a previous cytological study of another *Zygnema* strain that found 19
183 chromosomes (Prasad & Godward 1966). The total assembly size (71 megabases (Mb)) was close to
184 genome sizes estimated by flow cytometry, fluorescence staining (Feng et al., 2021), and k-mer frequency
185 analysis (**Figure S2, Table S1B**). The high mapping rates of UTEX 1559 and UTEX 1560 Illumina reads
186 to the SAG 698-1b genome (97.16% and 97.12%, respectively) confirm that these genomes are from

187 close relatives, in agreement with the strains' history (see **Supplemental Text 1**). We thus used
 188 chromatin conformation data from SAG 698-1b to scaffold UTEX 1559 and UTEX 1560 assemblies,
 189 which resulted in 20 chromosomes containing 97.3% and 98.3% of the total assemblies, respectively. The
 190 three new *Zygnema circumcarinatum* genomes represent the first chromosome-level assemblies for any
 191 streptophyte alga (**Table 1**).

192
 193 **Table 1:** Genome assembly statistics for the new *Zygnema* genomes and available streptophyte algae
 194 (Hori et al., 2014; Nishiyama et al., 2018; Cheng et al., 2019; Jiao et al., 2020; Wang et al., 2020) (see
 195 Table S1B-E for further details). Mapping rate of *Z. circumcarinatum* UTEX 1560 was calculated by
 196 using SAG 698-1b RNA-seq reads mapped to UTEX 1560 genome.

Species (strain)	Assembly size (Mb)	BUSCO (%)	N50 (kb)	Num. scaffolds (pseudochr.)	RNA-seq mapping rate (%)
<i>Zygnema circumcarinatum</i> SAG 698-1b	71.0	89.8	3958.3	90 (20)	97.2
<i>Zygnema circumcarinatum</i> UTEX 1559	71.3	88.2	3970.3	614 (20)	98.3
<i>Zygnema circumcarinatum</i> UTEX1560	67.3	87.9	3792.7	514 (20)	95.9**
<i>Zygnema cf. cylindricum</i> SAG 698-1a	359.8	70.6	213.9	3,587	88.3
<i>Mesotaenium endlicherianum</i> SAG 12.97	163	78.1	448.4	13,861	94.4
<i>Penium margaritaceum</i> SAG 2640	3661	49.8	116.2	332,786	96.8
<i>Spirogløea muscicola</i> CCAC 0214	174	84.7	566.4	17,449	95.2
<i>Chara braunii</i> S276	1430	78.0	2300	11,654	89.5
<i>Klebsormidium nitens</i> NIES-2285	104	94.9	134.9	1,814	98.1
<i>Chlorokybus melkonianii</i> CCAC 0220	74	93.3	752.4	3,809	96.5
<i>Mesostigma viride</i> CCAC 1140	281	59.2	113.2	6,924	84.3

197
 198 The plastome (157,548 base pairs (bp)) and mitogenome (216,190 bp, **Figure S3**) of SAG 698-1b
 199 were assembled into complete circular genomes (**Figure S3**). The plastome of SAG 698-1b is identical to
 200 those of UTEX 1559 (GenBank ID MT040697; Orton et al., 2020) and UTEX 1560. The mitogenomes of
 201 SAG 698-1b (OQ319605) and UTEX 1560 are identical but slightly longer than that of UTEX 1559
 202 (MT040698, 215,954 bp; Orton et al., 2020) due to extra repeats (**Figure S4**). These observations agree
 203 with the history of the strains (see Supplementary Information).

204 The nuclear genome assembly of SAG 698-1a is four times larger (360 Mb) than those of *Z.*
 205 *circumcarinatum* (**Table 1, Figure 1E**). K-mer frequencies (**Figure S2**) strongly suggest that SAG 698-
 206 1a is a diploid organism with an estimated heterozygosity rate of 2.22%. This supports previous reports of
 207 frequent polyploidy in Zygnematophyceae (Allen, 1958). The marked genome size differences further
 208 support the notion that they are two different species (Table 1, pictures in **Figure 1A**). Following a recent
 209 study (Feng et al. 2021), we refer to SAG 698-1a as *Zygnema cf. cylindricum*. Our molecular clock
 210 analyses suggest that *Z. cf. cylindricum* (SAG 698-1a) and *Z. circumcarinatum* (SAG 698-1b, UTEX
 211 1559, UTEX 1560) diverged from one another around 236 Ma (**Figure 1E, Table S1F**).

212 The plastome of SAG 698-1a was available (Turmel et al., 2005) and we here determined its
 213 mitogenome (OQ316644) (**Figure S5**), which, at 323,370 bp in size, is the largest known among
 214 streptophyte algae. Compared to SAG 698-1b, the mitogenome of SAG 698-1a contains more and much
 215 longer introns (**Table S1G,H**).

216

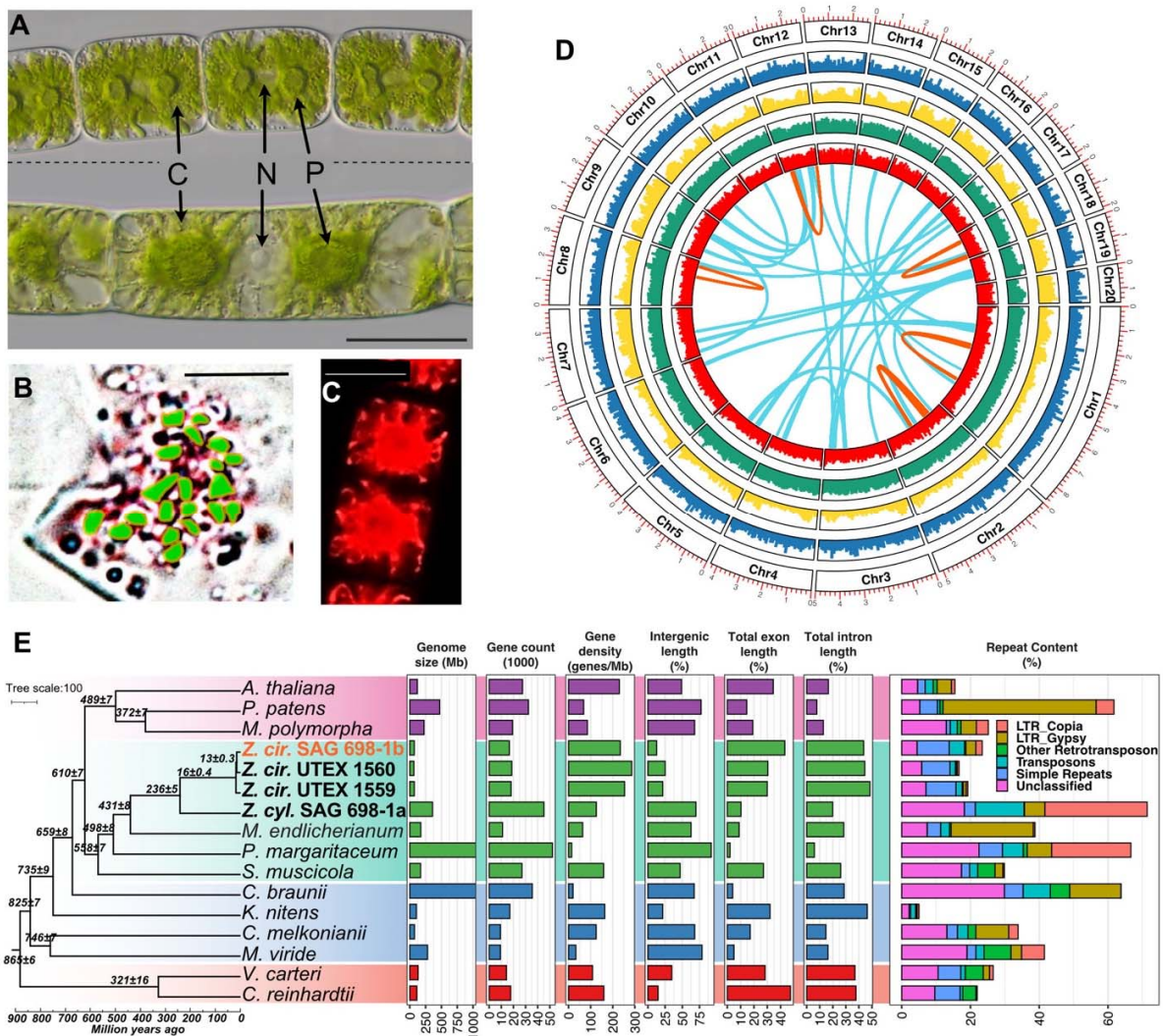


Figure 1: (A) Three cells of a vegetative filament of SAG 698-1b (top) compared to one cell of a vegetative filament of SAG 698-1a (bottom, both samples 1 month old). Scale bar: 20 μ m; C chloroplast; N nucleus; P pyrenoid. One-cell filament contains two chloroplasts and one nucleus. (B) Chromosome counting on light micrographs of SAG 698-1b fixed and stained with acetocarmine at prophase (0.5 months old); count was also performed in metaphase and telophase. Green dots represent the 20 chromosomes which were counted after rendering a stack of \sim 100 images (scale bar: 10 μ m); see Figure S1 for the original images. (C) Confocal laser scanning image of one SAG 698-1b cell (0.5 month). Scale bar: 20 μ m. (D) Chromosome-level assembly of SAG 698-1b genome. Concentric rings show chromosome (Chr) numbers, gene density (blue), repeat density (yellow), RNA-seq mapping density (\log_{10} (FPKM)) (dark green), and GC% density (red). Red and green links show respectively intra- and inter-chromosomal syntenic blocks. (E) Comparison of genome properties for 13 algal and 3 land plant species. The time-calibrated species tree was built from 493 low-copy genes (all nodes supported by $>97\%$ non-parametric bootstrap; numbers at nodes are estimated divergence times (mean \pm standard deviation) (see **Table S1F** for details). Data for bar plot can be found in **Table S1I,J**.

***Z. circumcarinatum* has the smallest sequenced streptophyte algal genome and no recent whole genome duplications**

The three *Z. circumcarinatum* genomes reported here have the smallest nuclear genomes of all streptophyte algae sequenced thus far (**Table 1**). They have the highest protein coding gene density, smallest percentage of intergenic regions, highest exon percentage, and lowest repeat content in

236 Zygnematophyceae (**Figure 1E, Table S1I**). The genome of SAG 698-1b contains 23.4% of repeats
237 (**Table S1J**), mostly consisting of simple repeats (6.4%) and transposable elements of the MITE (4.3%),
238 Gypsy (2.9%), and Copia (1.9%) families. For comparison, the *Zygnema cf. cylindricum* SAG 698-1a
239 genome contains 73.3% of repeats, consisting of Copia (29.8%), MITE (11.6%), Gypsy (5.9%), and
240 simple repeats (2.1%). The phylogenetic position and genome sizes of *Z. circumcarinatum* suggest
241 genomic streamlining in this species, as shown also for *K. nitens* (Hori et al. 2014) and perhaps *C.*
242 *melkonianii* (Wang et al., 2020). While the mechanisms of genomic streamlining are obscure, it clearly
243 occurred independently in different streptophytes; genome shrinkage may be a signature of adaptations to
244 new ecological niches (Bhattacharya et al., 2018).

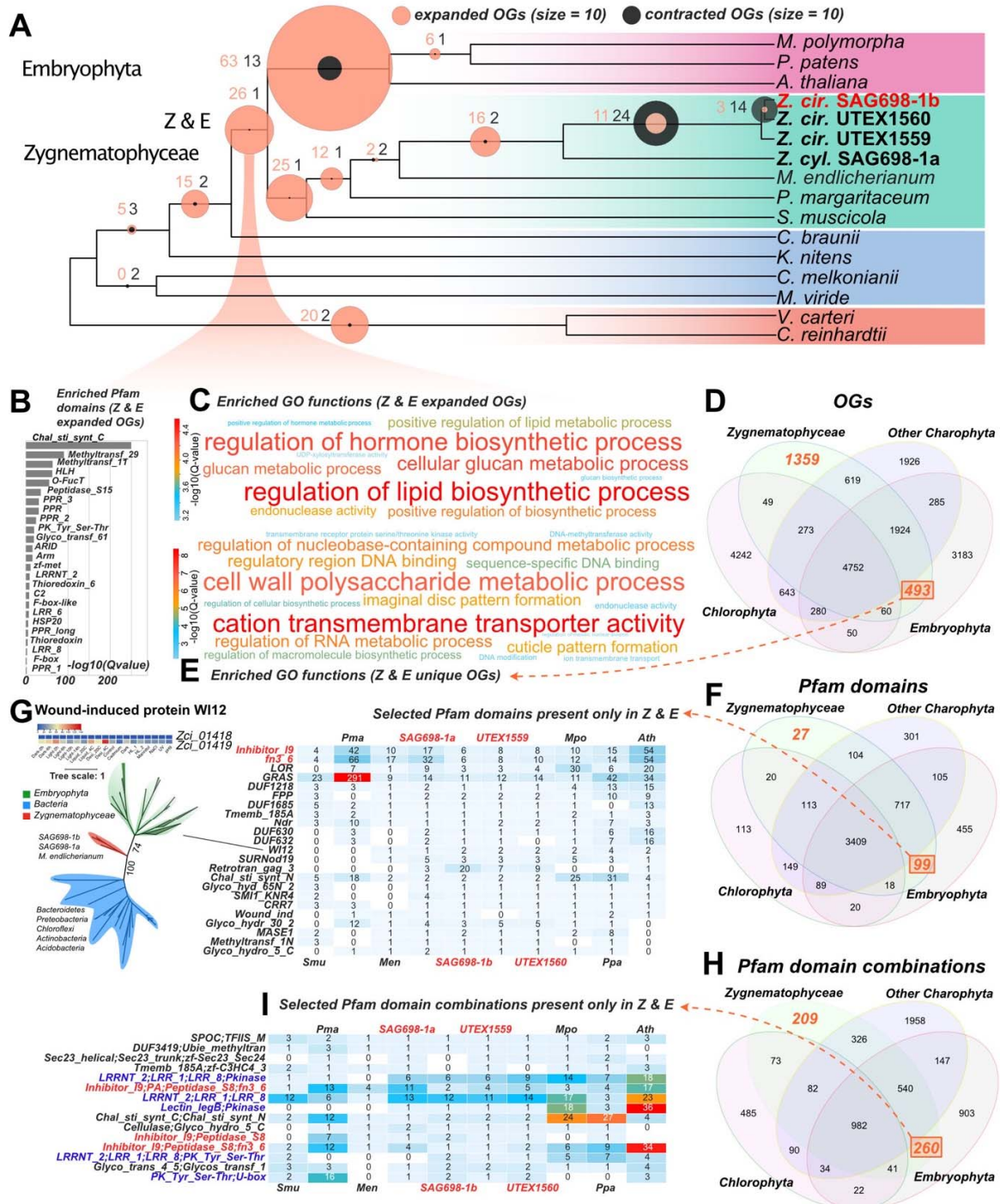
245 No evidence for whole genome duplication (WGD) was found in SAG 698-1b. MCscan (Wang et al.,
246 2012) was used to identify regions of conserved synteny, finding 28 syntenic blocks (≥ 4 genes per block,
247 distance between two colinear regions < 20 genes) totaling 236 genes (1.44% of the 16,617 annotated
248 genes) (**Figure S6A**). Increasing the distance among blocks to < 30 genes identified 190 syntenic blocks
249 containing 1,298 (7.9%) genes (**Figure S6B**). The distribution of synonymous distances among
250 paralogous gene pairs (Ks) in syntenic blocks found a single peak at Ks ~ 0.2 (**Figure S6C**). This agrees
251 with the apparent lack of WGD in *Z. circumcarinatum*. Because chromosome-level assemblies are so far
252 lacking for other streptophyte algae, similar analyses were conducted for the moss *P. patens*, reported to
253 have had at least two WGDs (Lang et al., 2018; Rensing et al., 2008). *P. patens* had more syntenic blocks
254 (**Figure S6D: 20% genes in syntenic blocks and S6E: 24% genes in syntenic blocks**), and a single peak
255 Ks ~ 0.8 (**Figure S6F**) that is likely a fusion of the two known WGDs (Ks ~ 0.5 – 0.65 and Ks ~ 0.75 – 0.9
256 (Lang et al., 2018)). In the absence of evidence for WGD in SAG 698-1b, synteny blocks are likely the
257 result from segmental duplications. Congruently, the top enriched Pfam domains in these regions are
258 related to retrotransposons.

259

260 **Comparisons of the three *Z. circumcarinatum* genomes**

261 Our phylogenetic analyses show that SAG 698-1b and UTEX 1560 are closer to each other than to UTEX
262 1559 (**Figure 1E**), a result confirmed by seven- and four-genome phylogenies using thousands of single
263 copy genes (**Figure S7**). Gauch (1966) reported that UTEX 1559 was a non-functional mating type (+)
264 whereas UTEX 1560 and SAG 698-1b were functional mating type (-). This agrees with our conjugation
265 experiments that failed to conjugate UTEX 1559 with UTEX 1560 or SAG 698-1b. Whole genome
266 alignments of UTEX 1559 and UTEX 1560 against SAG 698-1b respectively, identified high alignment
267 coverages and regions of conserved synteny across most chromosomes (**Figure S8A,B**). SAG 698-1b
268 showed higher alignment coverage (less genomic rearrangements) with UTEX 1560 than to UTEX 1559,
269 although aligned regions are slightly more similar to UTEX 1559 (**Figure S8C,D**). Among the three *Z.*
270 *circumcarinatum* genomes, chromosomes 20, 13, and 16 differ the most (see **Figure S8E** for a three-way
271 alignment of Chr20), suggesting that they might contain sex/mating determination loci. The mating loci in
272 *Zygnema* are so far unknown and we did not identify homologs of the sex hormone proteins (PR-IP and
273 its inducer) described in *Closterium* (but homologs were found in *Penium margaritaceum*) (**Table S1K**).
274 Gene content comparison found 17,644 core genes, i.e., shared by all three *Z. circumcarinatum* genomes
275 (**Figure S8F**). Most of the unique and shell genes have no known Pfam domains and no hits in NR (NCBI
276 non-redundant protein sequence database) and thus remain functionally unknown.

277



278
279
280
281
282
283
284
285

Figure 2: Comparative genomics of 13 algal and 3 land plant genomes. (A) Gene family expansion and contraction patterns estimated by CAFE using Orthofinder-identified orthogroups and the time-calibrated phylogeny of Figure 1E. Key nodes are indicated on the tree and circles denote significant expansions and contractions (circle size reflects the number of expanded/contracted orthogroups/OGs). (B) Pfam domain enrichment for genes on the node leading to Zygnematophyceae and Embryophyta (Z&E). (C) Functional (GO) enrichment for the Z&E node. (D) Orthogroups overlap among Chlorophyta, Embryophyta, Zygnematophyceae, and other streptophyte algae. (E) Enriched GO terms in the 493 orthogroups exclusive to Zygnematophyceae and Embryophyta. (F) Pfam domain overlap among Chlorophyta, Embryophyta, Zygnematophyceae, and

286 other streptophyte algae. (G) Exclusive Pfam domains found only in Zygnematophyceae and Embryophyta. One Pfam family
287 WI12 was studied with phylogenetic analysis, suggesting a possible horizontal gene transfer from bacteria and expression
288 response to stresses. (H) Pfam domain combination overlap among Chlorophyta, Embryophyta, Zygnematophyceae, and other
289 streptophyte algae. (I) Exclusive Pfam domain combinations in Zygnematophyceae and Embryophyta.

291 **Comparative genomics identifies significantly enriched orthogroups and domains in** 292 **Zygnematophyceae and land plants**

293 Comparative genomics were performed with annotated proteins from 16 representative algal and plant
294 genomes, clustered into orthogroups (groups of orthologs or OGs) by OrthoFinder. 4,752 orthogroups
295 contained at least one representative of Chlorophyta, Embryophyta, Zygnematophyceae, and other
296 streptophyte algae (**Figure 2A,B,C,D**). We identified clade-specific orthogroups according to the species
297 phylogeny (**Figure 2A**); the enrichments in gene ontology (GO) terms (**Figure 2C**) and Pfam domains
298 (**Figure 2B**) were inferred by binomial test q-values against the background of GO terms and Pfam
299 domains present in the general set of 4,752 orthogroups (**Figure 2D**). In the ancestor of
300 Zygnematophyceae and Embryophyta, we inferred an overrepresentation of Pfam domains (**Figure 2B**)
301 including (i) Chal_sti_synt_C (found in the key enzyme of the flavonoid pathway chalcone synthase
302 (CHS)), (ii) Methyltransf_29 (found in an *Arabidopsis thaliana* gene (AT1G19430) required for cell
303 adhesion (Krupková et al. 2007)), (iii) PPR domains involved in organellar RNA binding and editing, and
304 (iv) domains related to plant immunity such as LRR and Peptidase_S15 (Muszewska et al., 2017),
305 PK_Tyr_Ser-Thr, and Thioredoxins (Kumari et al., 2021). Some of these orthogroups and domains
306 overrepresented in Zygnematophyceae + Embryophyta could be the results of horizontal gene transfer
307 (HGT), such as Chal_sti_synt_C (Ma et al., 2022) and O-FucT (**Figure S9**). O-FucT is the GDP-fucose
308 protein O-fucosyltransferase domain (PF10250), and its homologs in streptophytes are likely transferred
309 from arbuscular mycorrhizal (AM) fungi Mucoromycotina and Glomeromycota (**Figure S9**). In SAG
310 698-1b, the O-FucT domain-containing gene (Zci_09922) is expressed under drought and cold stresses.

311 The Zygnematophyceae + Embryophyta (Z+E) ancestor showed enriched GO terms related to
312 biosynthesis of phytohormones, lipids, and glucan (**Figure 2C**). 493 orthogroups were exclusively found in
313 Z+E and not in any other studied species (**Figure 2D**), and these were enriched in GO terms “cation
314 transmembrane transporter activity” and “cell wall polysaccharide metabolic process”. 1,359 orthogroups
315 were Zygnematophyceae-specific, with enriched GO terms “phosphorylation”, “pyrophosphatase
316 activity”, “transmembrane receptor protein serine/threonine kinase activity”, “cellular response to abscisic
317 acid stimulus”, and “polysaccharide biosynthetic process” (**Figure 2E**).

318 Regarding Pfam domains, we found 3,409 to be present in at least one representative of
319 Chlorophyta, Embryophyta, Zygnematophyceae, and other streptophyte algae, whereas 99 were
320 exclusive to Z+E, and 27 were Zygnematophyceae-specific (**Figure 2F**). Unlike the Pfam domains
321 enriched in the Z+E ancestor (**Figure 2B**), the 99 or 27 domains were not found outside of
322 Zygnematophyceae and Embryophyta. The most prevalent domains that are exclusively of Z+E (**Figure**
323 **2G**) include several transcription factors (more details below). In some cases, the domains exclusive to
324 Z+E could be the result of HGT. For example, the Inhibitor_I9 and fn3_6 domains are among the most
325 abundant in Z+E (**Figure 2G**) and often co-exist with Peptidase_S8 domain in plant subtilases (SBTs)
326 (**Figure 2I**), which has been reported to originate by HGT from bacteria (Xu et al., 2019). The WI12
327 domain, named after the plant wound-induced cell wall protein WI12, was also possibly gained by HGT
328 from bacteria (**Figure 2G**). WI12 expression is induced by wounding, salt, methyl jasmonate, and
329 pathogen infection in the facultative halophyte ice plant (Yen et al., 2001), and plays significant roles in
330 reinforcing cell wall and is involved in the defense to cyst nematodes in soybean (Dong and Hudson,

331 2022). In SAG 698-1b, one of the two genes containing the WI12 domain (Zci_01419) is significantly
332 upregulated under desiccation and cold stresses.

333 Because new functions can arise through domain combinations, we searched for lineage-specific
334 Pfam domain combinations in our dataset (**Figure 2H**). 982 Pfam domain combinations are shared by all
335 studied genomes, 260 being unique to Z+E and 209 to Zygnematophyceae. Among those exclusive to
336 Z+E (**Figure 2I**), we found Lectin_legB and Pkinase domains, which despite having older evolutionary
337 origins, were only combined into a protein in the Z+E ancestor (e.g., Zci_10218). A search for the 99 and
338 260 domain combinations by BLASTP against the NR database (**Table S2A,B**) failed to identify such
339 combinations in sequenced chlorophyte or streptophyte algal genomes but were occasionally found in
340 fungi, prokaryotes, and viruses. This pattern could be the product of HGT, functional convergence,
341 rampant gene loss, or contamination. Combining existing protein domains is a powerful mechanism for
342 functional innovation, as shown for, cell adhesion, cell communication and differentiation (Itoh et al.,
343 2007; Vogel et al., 2004).

344

345 **Orthogroup expansions reveal increased sophistication and resilience**

346 We inferred 26 significantly expanded orthogroups in the Z+E ancestor (**Figure 2A; Table S3A,B**).
347 Three are related to phytohormone signaling: ethylene-responsive element (ERE)-binding factors (OG 22)
348 and PP2Cs (OG 548 and OG 830), regulatory hubs for diverse responses, including to ABA, whose
349 evolutionary origin—albeit with an ABA-independent role in algae—was already pinpointed (de Vries et
350 al., 2018; Sun et al., 2019; Cheng et al., 2019). Several expansions suggest more sophisticated gene
351 networks featuring calcium signaling (putative calcium-dependent protein kinases; OG 19) and ubiquitin-
352 mediated proteolysis (RING/U-box proteins; OG 78), both cornerstones in plant stress response and
353 environmental signaling (Yee and Goring, 2009; Reddy et al., 2012). We inferred expansion of
354 transmembrane transporters such as major facilitator superfamily proteins (OG 169), amino acid
355 transporters (OG 333), acetate channels (OG 353), RAB GTPases (OG 174), and Golgi SNAREs (OG
356 963). Two orthogroups might be involved in interactions with microbes and fungi, including subtilases
357 (OG 23; Xu et al. 2019) and glycosyl hydrolases with a chitin domain (OG 857; (Parrent et al. 2009)).
358 Growth and development are underpinned by expanded beta-glucosidases (OG 85) involved in
359 xyloglucan biosynthesis and/or plant chemical defense (Morant et al. 2008), and developmental regulators
360 root hair defective 3 GTP-binding proteins (OG 996) (Yuen et al., 2005) and ARID/BRIGHT DNA
361 binding TFs (transcription factors) (OG 2169). A comprehensive analysis of transcription associated
362 proteins (TAPs) with TAPscan v.3 (Petroll et al. 2021) showed higher numbers of TFs in land plants as
363 compared to algae, as expected due to their more complex bodies (Fig. 5B; Table S3B). *Zygnema* species
364 had comparatively more TAPs than other algae (525-706 vs. 269 in *Ulva* and 371 in *Chlorokybus*; Table
365 5B; Table S3B). All four studied *Zygnema* strains show similar TAP profiles, with the exception of
366 MYB-related TF family and PHD (plant homeodomain), which were, respectively, 2- and ≥ 9 -fold larger
367 in SAG 698-1a.

368 The common ancestor of Zygnematophyceae displayed 25 significantly expanded orthogroups
369 (**Figure 2A; Table S3B**). Most expanded are alpha-fucosyltransferases (OG 89) involved in xyloglucan
370 fucosylation (Faik et al., 2000). We found ethylene sensors and histidine kinase-containing proteins (OG
371 94), bolstering the idea that two-component signaling is important and active in filamentous
372 Zygnematophyceae (Ju et al., 2015; Bowman et al., 2019). Several orthogroups were associated with
373 typical terrestrial stressors: aldo-keto reductases closely related to *M. polymorpha* Mp2g01000 (OG 942)
374 that could reflect ROS scavenging machinery (Stiti et al. 2021), DNA helicases for DNA repair and

375 recombination (OG 1471), and methyltransferases (OG 269, OG 369) that could underpin specialized
376 metabolism of phenylpropanoids for stress response (Lam et al., 2007). The phenylpropanoid pathway is
377 in fact a well-known response to terrestrial stressors (Dixon and Paiva, 1995) and its enzymes have deep
378 roots in streptophyte evolution (de Vries et al., 2021; Rieseberg et al., 2023); *Zygnema* has a set of
379 phenylpropanoid enzyme homologs comparable to those reported before (**Figure S14**). We found
380 expansions in putative light-oxygen-voltage sensitive (LOV)-domain containing proteins (OG 1897),
381 photoreceptors mediating responses to environmental cues (Glantz et al., 2016) that imply a more
382 elaborate response to rapidly changing light regimes typical for terrestrial habitats (see also below).
383 Several expanded orthogroups relate to developmental processes. Expanded signaling and transport,
384 possibly related to filamentous growth, include calcium signaling (OG 56), zinc-induced facilitators (OG
385 258), cysteine-rich fibroblast growth factor receptors found in the Golgi apparatus (OG 518), and
386 cation/H⁺ antiporters (OG 809). Cation/H⁺ antiporters are closely related to *A. thaliana* nhx5/nhx6 that
387 act on pH and ion homeostasis in the endosome and are key for membrane trafficking in the trans-Golgi
388 network (Bassil et al., 2011; McKay et al., 2022), and diverse developmental processes (Dragwidge et al.,
389 2018). A dynein homolog (OG 72) was significantly contracted in Zygnematophyceae (**Table S3B**),
390 which might be associated with cytokinesis of cilia and flagella and thus in line with the loss of motile
391 gametes in Zygnematophyceae and Embryophyta (OG 72 was also identified as contracted in that
392 ancestor).

393 *Zygnema* is noteworthy among algae for its stress resilience and can grow in extreme habitats such as
394 the Arctic, where it is abundant (Pichrtová et al., 2018; Rippin et al., 2019). Its broad ecological
395 amplitude is confirmed by numerous studies on temperature and light stress (summarized in Permann et
396 al., 2022b) and desiccation stress (Becker et al., 2020). The common ancestor of all four *Zygnema* strains
397 is inferred to have had 16 significantly expanded orthogroups (**Figure 2A; Table S3B**), including PP2Cs
398 (OG 548), early light-inducible proteins (ELIPs; OG 97), and low-CO₂ inducible proteins (LCICs, OG
399 459). Previous studies found ELIPs to be among the top upregulated genes when *Zygnema* is exposed to
400 environmental challenges, including growth in the Arctic (Rippin et al., 2017; 2019; de Vries et al., 2018).
401 We further found expansion of HSP70 chaperones (OG 277) involved in protein folding and stress
402 response as previously observed in *Mougeotia* and *Spirogyra* (de Vries et al., 2020), as well as Leucine-
403 rich repeat (LRR) proteins (OG 35 and OG 995) related to plant-microbe interactions. Two orthogroups
404 were significantly contracted: GTP binding elongation factor *Tu* family (OG 251) and seven
405 transmembrane MLO family protein (OG 320). On balance, the evolution of gene families reflects
406 *Zygnema*'s resilience in the face of challenging habitats.

407

408 **Gene gains facilitated major cell wall innovations**

409 The earliest land plants had to overcome a wide range of stressors (Fürst-Jansen et al., 2020) and cell
410 walls are the first layer of protection from the environment. We reconstructed the evolutionary history of
411 38 cell wall-related enzyme families (**Table S1L**). Large gene families were split into 76 well-supported
412 subfamilies (>70% non-parametric bootstrap supports) according to phylogenetic trees (Data S1, homolog
413 counts are in **Table S1M** and **Figure 3A**). Most subfamilies belong to carbohydrate active enzyme
414 (CAZyme) families known for the synthesis and modifications of celluloses, xyloglucans, mixed-linkage
415 glucans, mannans, xylans, arabinogalactan proteins (AGPs), and pectins (**Table S1L,M, Figure 3A**).
416 CAZymes include glycosyl transferases (GTs), glycosyl hydrolases (GHs), carbohydrate esterases (CEs),
417 and polysaccharide lyases (PLs). Analyzing the 76 enzyme subfamilies revealed that (i) Z+E share all the
418 major enzymes for the synthesis and modifications of the diverse polysaccharide components, including

419 those for sidechains and modifications (**Figure 3B**, 42-56 subfamilies in Zygnematophyceae vs 63-69 in
420 Embryophyta); (ii) many of the enzymes for cell wall innovations, especially for polysaccharide backbone
421 synthesis, have older evolutionary origins in the common ancestor of Klebsormidiophyceae,
422 Charophyceae, Zygnematophyceae, and Embryophyta (**Figure 3B**, 34-69 subfamilies vs 5-7 in
423 Chlorokybophyceae and Mesostigmatophyceae). Many of such subfamilies are expanded in
424 Zygnematophyceae (**Figure 3B % of genes**, e.g., GH16_20, GT77, CE8, CE13 in **Figure 3A**); (iii) genes
425 involved in the syntheses of different cell wall polymers (backbones and sidechains) are co-expressed in
426 SAG 698-1b (**Figure 3C**); (iv) phylogenetic patterns (**Data S1**) suggest that HGT could have played an
427 important role in the origin of the enzymatic toolbox for cell wall polysaccharide metabolism (**Figure 3A**,
428 **Supplemental Text 2**). HGT is more common for degradation enzymes (e.g., GH5_7, GH16_20,
429 GH43_24, GH95, GH27, GH30_5, GH79, GH28, PL1, PL4) but it is also observed for GT enzymes; (v)
430 frequent gene loss creates scattered distribution of homologs in Streptophyta (**Figure 3A**; e.g., *Zygnema*
431 lacks GH5_7, GH35, GT29, GT8, CE8, GH28, PL1).

432 We performed a careful analysis of the GT2 family, which contains major cell wall synthesis
433 enzymes such as cellulose synthase (CesA) for the synthesis of cellulose and Csl (CesA-like) for
434 hemicellulose backbones (**Figure 3D**). Among the 11 SAG 698-1b CesA/Csl homologs, ZcCesA1
435 (Zci_04468), ZcCslL1 (Zci_07893), ZcCslC (Zci_01359), ZcCslN (Zci_08939), ZcCslP1 (Zci_0910) are
436 highly expressed in response to various stresses (**Figure 3E**), in agreement with Fitzek et al (2019) who
437 showed response to osmotic stress. The two CesAs homologs in SAG 698-1b (**Figure 4D, 4F, Data S1-1**)
438 and all other Zygnematophyceae homologs are orthologs of land plant CesA; Zygnematophyceae also
439 have a second CesA homolog (ZcCesA2; Zci_03055) not found in land plants but shared with other
440 streptophytes. The presence of land plant-like CesAs only in Zygnematophyceae suggests that the CSC
441 (cellulose synthase complex) structure in a six-subunit rosette typical of land plants evolved in their
442 common ancestor; this agrees with observations made by electron microscopy and isotope labeling that
443 found an hexameric rosette CSC in Zygnematophyceae but no in other algae (Tsekos, 1999). ZcCesA1
444 (Zci_04468) is co-expressed with four known plant primary cell wall CSC component core genes: KOR
445 (Zci_10931), CC1 (Zci_04753), CSI1 (Zci_02943), THE (Zci_09278) (**Figure 3C**). This extends
446 previous observations (Lampugnani et al., 2021) suggesting that co-expression of CSC component genes
447 is evolutionarily conserved since the common ancestor of Zygnematophyceae and land plants.

470 gained by an ancient HGT from bacteria into the common ancestor of Streptophyta algae (**Figure 3F**).
471 CIsK was thought to be restricted to Chlorophyta (Yin et al., 2009; Yin et al., 2014) but has recently been
472 found in Zygnematophyceae (Mikkelsen et al., 2021). CslL is a new Zygnematophyceae-specific family.
473 CslK and CslL might be responsible for the mannan backbone synthesis in green algae (Chlorophyta and
474 Zygnematophyceae) given their closer relationship to CslA than CslC (**Data S1-1, Figure 3D**). CslO, also
475 found in non-seed land plants (**Figure 3G**), contains a *Physcomitrium pattens* homolog (Pp3c12_24670)
476 involved in the synthesis of the *P. patens*-specific polysaccharide arabinoglucan (AGlc) (Roberts et al.,
477 2018). CslO is closely related to a large fungal clade containing Tft1 (XP_748682.1) characterized for the
478 synthesis of MLG, a fungal cell wall component (Samar et al., 2015). This opens the possibility that CslO
479 was horizontally transferred from fungi (**Figure 3G**). CslP is absent in land plants, and together with
480 CslO further clustered with known CesAs from oomycetes, tunicates, amoeba, and red algae (Blanton et
481 al., 2000; Blum et al., 2010; Matthews et al., 2010; Matthysse et al., 2004) (**Figure 3G**). The close
482 affinity with a large bacterial clade of BcsA/CesAs (>10,000 proteins from >10 phyla; BLASTP of
483 ZcCslP1 against NR, E-value < 1e-30), including experimentally characterized cyanobacterial CesAs,
484 suggests that eukaryotic CslP-like proteins could have originated via HGT (**Figure 3G, Data S1-12**).
485 CslO evolved from CslP-like proteins and was subjected to duplication and functional diversification in
486 eukaryotes (e.g., MLG synthesis in fungi and likely also in various microalgae, AGlc synthesis in
487 mosses), while the CesA function of CslP-like proteins is conserved in oomycetes, tunicates, amoeba, red
488 algae, and possibly also in green algae with CslP. CslN homologs are restricted to Zygnematophyceae and
489 bacteria (no significant NCBI NR hits outside them at E-value < 1e-40), which suggests a possible HGT
490 from bacterial CesA or other beta-glucan synthases (**Figure 3H, Data S1-13**) that often contain N-
491 terminal GGDEF and REC domains. The bacterial BcsA-like CslN and CslP, if biochemically
492 characterized as CesAs in Zygnematophyceae in the future, are clearly of a distinct origin than land plant
493 CesAs, thus representing an example of convergent evolution.

494 Fucosyltransferase (FUT, member of the GT37 family), is significantly expanded in
495 Zygnematophyceae (**Data S1-49, OG 89** is the most expanded orthogroup; see below). Xyloglucan
496 fucosylation is implicated in stress response and was long thought to be land-plant-specific but was
497 recently shown to be present in the Zygnematophyceae *Mesotaenium caldariorum* (Tryfona et al., 2014;
498 Mikkelsen et al., 2021). Our analyses show a sparse phylogenetic distribution of FUT homologs in
499 *Zygnema* and embryophytes and distant homologs in *Chara* and *Klebsormidium*, as well as in many soil
500 saprotroph Mortierellaceae fungi (Telagathoti et al., 2021) (**Data S1-19**). Phylogenetic patterns could be
501 compatible with an ancient HGT between plants and fungi (**Data S1-19**). The enzyme XTH (GH16_20),
502 which degrades the xyloglucan backbone, shows a similar pattern compatible with HGT between fungi
503 and streptophytes (**Data S1-20, S1-21**), in agreement with a recent report (Shinohara and Nishitani,
504 2021). Type II fucosidase (AXY8, GH95), involved in degrading fucosylated xyloglucans and
505 arabinogalactan proteins (Léonard et al., 2008; Wu et al., 2010), might have been acquired by HGT from
506 bacteria (**Data S1-22**). Overall, these findings agree with an earlier origin of xyloglucan backbone
507 enzymes and a later origin of side chain modification enzymes in the common ancestor of
508 Zygnematophyceae and Embryophyta (Del-Bem, 2018; Mikkelsen et al., 2021). Recent reports have
509 investigated the occurrence of homogalacturonans (e.g., Herbruger et al. 2019) and AGPs in *Zygnema*
510 (Palacio-Lopez et al. 2019) and the filamentous Zygnematophyceae *Spirogyra*, where recently a
511 rhamnogalactan protein has been described (Pfeifer et al., 2022); further, the specific cell wall
512 composition of zygospores of the filamentous Zygnematophyceae *Mougeotia* and *Spirogyra* have been
513 described (Permann et al., 2021; 2022a). Cell wall modifications by endotransglycosylases and novel

514 transglycosylation activities between xyloglucan and xylan, xyloglucan and galactomannan were
515 described previously (Herburger et al., 2018). Overall, the phylogenetic analyses of key cell wall enzymes
516 (**Data S1, Supplemental Text 2, Table S1L**) highlighted the importance of ancient HGTs contributing to
517 evolutionary innovations of cell walls, similarly to what has been proposed for other traits (Yue et al.,
518 2012; Cheng et al., 2019; Ma et al., 2022).

519

520 **Gene co-expression networks in *Zygnema circumcarinatum* SAG 698-1b**

521 We explored functional gene modules in *Zygnema circumcarinatum* (SAG 698-1b) by inferring gene co-
522 expression networks from RNA-seq data from a diverse set of 19 growth conditions (various day-light
523 cycles and dark, liquid and agar cultures at 20°C, pH=9 and terrestrial stressors including cold treatment
524 at 4°C, desiccation at 4°C, high light, high light at 4°C, or UV; see Methods). We obtained 406 clusters
525 containing 17,881 out of the 20,030 gene isoforms annotated in the genome. Gene co-expression
526 networks are available through the CoNekT web portal (Proost and Mutwil, 2018;
527 <https://zygnema.sbs.ntu.edu.sg/>). We searched for homologs of genes related to (i) cell division and
528 development, (ii) multicellularity, (iii) stress response, (iv) transporters, (v) phytohormones, (vi) calcium
529 signaling, and (vii) plant-microbe interaction. Candidate genes were drawn from the literature and the
530 expanded orthogroups. 150 out of 406 clusters showed co-occurrence of at least two such functional
531 categories, the most frequent co-occurrence being plant-microbe interaction and calcium signaling,
532 followed by plant-microbe interaction and stress (Figure 4A).

533

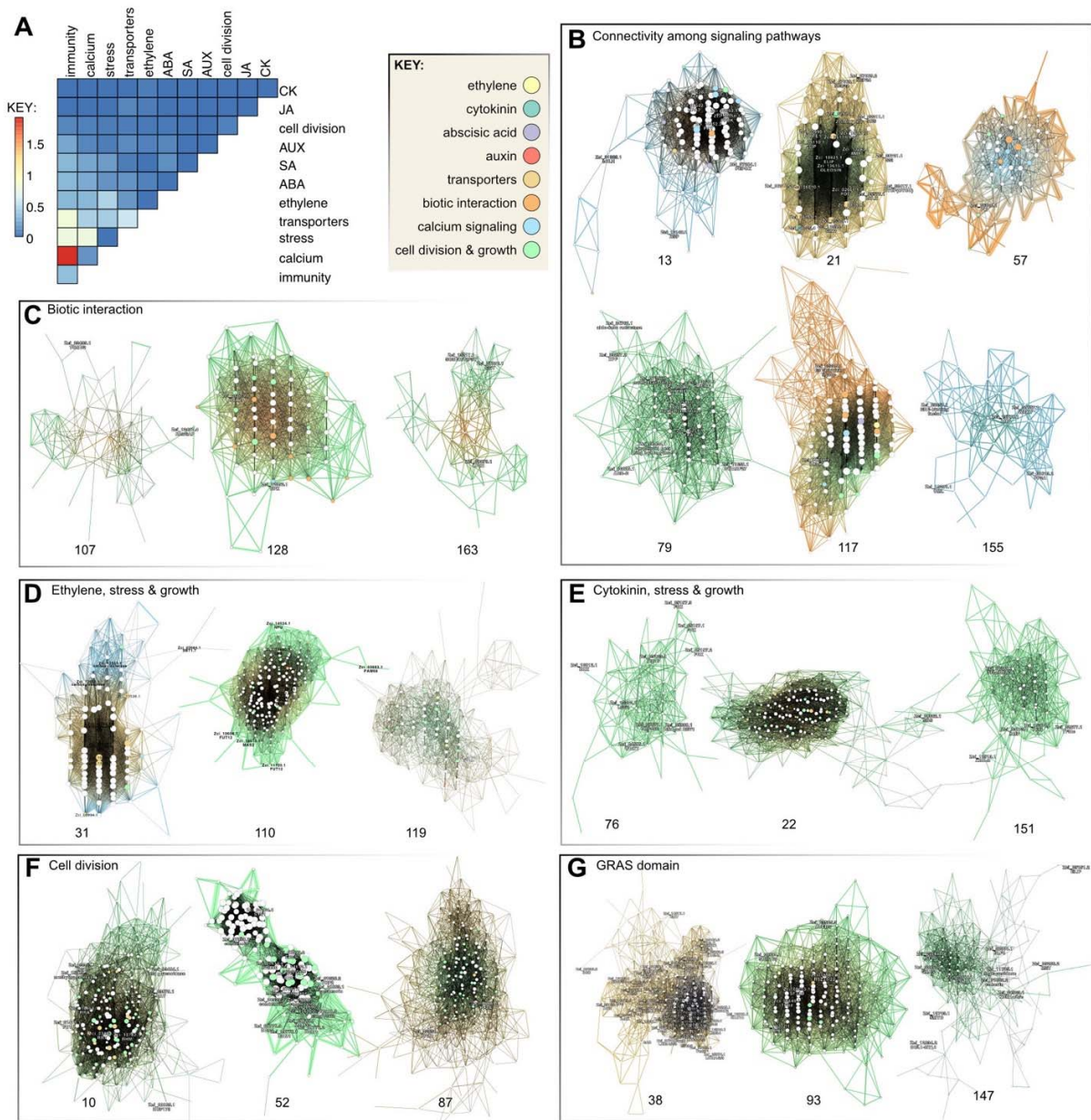
534 **On the deep evolutionary roots of the plant perceptron**

535 Land plants are sessile. Sensing environmental conditions and modulating growth alongside mounting
536 appropriate physiological responses is vital for plants, especially under adverse conditions. Several of our
537 co-expression clusters reveal the concerted action of diverse signaling pathways to sense, process, and
538 respond to environmental cues. This agrees with the concept of the plant perceptron (Scheres & van der
539 Putten 2017) that views plant biology as a molecular information-processing network—formed by genes
540 and their biochemical interactions—that enable adequate response given a combination of input cues.
541 This implies a high degree of connectivity among signaling pathways to modulate a response to highly
542 complex biotic and abiotic signals. Such connectivity establishes the foundation for an adaptive advantage
543 of multicellular morphogenesis, where cell differentiation can be fine-tuned for acclimation to
544 environmental cues.

545 In SAG 698-1b, many gene modules reflect a complex connectivity among signaling pathways
546 including phytohormones, calcium signaling, transporters, and cell division and developmental genes.
547 This shows that the plant perceptron has deeper evolutionary roots dating back at least to the common
548 ancestor of land plants and Zygnematophyceae. Interestingly, even though some phytohormones are
549 likely not present in *Zygnema* (see below), we observe gene homologs of their biosynthesis and signaling
550 co-expressing with well-known effectors, thus suggesting these genes were already involved in response
551 to abiotic and biotic stresses before the appearance of some phytohormones (see discussion in Fürst-
552 Jansen et al., 2020). Several clusters (**Figure 4B**) contain early light-induced proteins (ELIPs), light-
553 harvesting complex-related proteins that respond to light stimulus and can reduce photooxidative damage
554 by scavenging free chlorophyll (Hutin et al., 2003) under cold stress (cluster 21) —as shown for other
555 streptophyte algae (Han et al. 2013; de Vries et al. 2018)— but also under high light (cluster 20). In
556 cluster 20, we identified a heat shock protein and a PsbS homolog putatively involved in non-
557 photochemical quenching. Cluster 13 is deployed under dark conditions and shows expression of a

558 red/far-red light receptor phytochrome in the canonical PHY2 clade (Li et al., 2015; **Figure S15**).
559 Zygnematophyceae have a unique chimeric photoreceptor called neochrome, which is a combination of
560 red/far-red sensing phytochrome and blue-sensing phototropin (Suetsugu et al., 2005). Neochromes have
561 independently evolved twice, with a separate origin in hornworts (Li et al., 2014). In Zygnematophyceae,
562 neochrome is hypothesized to be involved in chloroplast rotation to maximize light reception and to avoid
563 photodamage (Suetsugu et al. 2005). In cluster 123, two neochrome homologs are co-expressed with
564 zeaxanthin epoxidase and a few photosynthesis-related genes, in agreement with its role in fine-tuning
565 photosynthesis.

566 Several clusters include leucine-rich repeat (LRR) proteins that might be involved in response to
567 biotic stresses. Downstream, late embryogenesis abundant (LEA) proteins and protein phosphatases type
568 2C (PP2C) are often found, plant protein groups involved in diverse stress responses (Dure et al. 1981).
569 Calcium signaling genes are also frequently expressed under stress conditions (cluster 117). Signal
570 transducers are expected to affect many downstream genes through deployment of transcription factors
571 (many visible in the clusters: WRKY, PSRP1, MYB, bHLH, mTERF, SCR), thereby allowing a tightly
572 orchestrated response in time and space. As expected from the involvement of several transcription
573 factors in cell division (e.g., MYB, SCR), we identify master cell cycle regulators (e.g., CYCP in cluster
574 21) and genes involved in cell division and development. Among phytohormone pathways, we highlight
575 the histidine phosphotransfer protein (HPT) putatively involved in cytokinin signaling (AHP homolog;
576 cluster 155), various ethylene responsive element-binding factors (clusters 57, 155), ABI1/2 homologs of
577 the ABA biosynthesis (cluster 21), and the PYL homolog (cluster 13) that is an ABA receptor in land
578 plants but probably not in algae (Sun et al., 2019). The expression of different ethylene response factors
579 under desiccation (cluster 57) or other stress conditions (cluster 155) reveals functional specialization. In
580 cluster 79, we identify ROS scavenging (cluster 79; Stiti et al. 2021; OG 942 expanded in
581 Zygnematophyceae), various genes of the carotenoid pathway, and DNA repair machinery. Under cold
582 conditions (cluster 21) and osmotic stress (cluster 103), oleosins are expressed for the formation of lipid
583 droplets. In line with the need to tightly control metabolism, various genes of the carbon metabolism are
584 also co-expressed, particularly under dark and cold conditions (clusters 13 and 21). Overall, the co-
585 expression data suggest a joint action of genes to sense the environment and modulate growth.



586
587
588
589
590
591
592
593
594
595
596
597
598

Figure 4. Gene co-expression modules in *Zynema circumcarinatum* SAG 698-1b. (A) Heatmap of per-module co-occurrence frequencies among genes associated with plant-microbe (p-m) interaction, calcium signaling, phytohormone, stress, transporters, cell division, and diverse phytohormones (ethylene, cytokinin, abscisic acid/ABA, auxin/AUX, jasmonic acid/JA, salicylic acid/SA); based on 150 out of 406 total modules showing co-occurrence of at least two functional categories. (B) Modules reflecting connectivity among signaling pathways, (C) biotic interaction, (D) ethylene, stress, and growth, (E) cytokinin, stress & growth, (F) cell division, (G) GRAS-domain containing genes. In gene networks, nodes are genes (size proportional to number of neighbors) and edges reflect co-expression (width proportional to Pearson's correlation coefficient and colors are those of interconnected genes); numbers below indicate network number; gradient colors of the edges highlight the two dominant gene categories indicated in the KEY. The full gene co-expression results can be accessed in our online portal (<https://zynema.sbs.ntu.edu.sg/>).

Gene modules associated with cell division and development

599 We compiled an extended list of 270 genes with experimental evidence for their involvement in mitosis
600 and cytokinesis in *Arabidopsis thaliana* (including cytoskeleton and endomembrane transport and
601 upstream regulators of cell division; Table S3D). Homolog distribution across streptophyte and
602 chlorophyte genomes revealed secondary loss in *Zygnema* of microtubule plus tip CLASP and SPIRAL1
603 genes key to microtubule dynamics. There are important changes in the AUGMIN complex involved in
604 microtubule nucleation, and GIP1 and GIP2 (γ -tubulin complex interactors) are missing from all analyzed
605 Zygnematophyceae genomes. Homologs of AUGMIN2 and EDE1-like show weak conservation in
606 Zygnematophyceae. Together, these data suggest that the Zygnematophyceae, and especially *Zygnema*,
607 exhibit major differences in microtubule dynamics.

608 Some cell division genes are land plant-specific, including TANGLED (see Nishiyama et al., 2018),
609 TRM (TON1-recruiting motif), SOSEKI, and EPF1 (epidermal patterning factor), which have pivotal
610 roles in cell polarity and division and may have underpinned the multicellularity and increased
611 morphological complexity of land plant. We also identified genes that likely originated in the common
612 ancestor of Z+E: UGT1, SUN1, SUN2, and LONESOME HIGHWAY. These genes did not have
613 reciprocal best BLAST hits in other streptophyte algae, although likely paralogs were found in
614 *Klebsormidium nitens* (Table S3D). The clearest cases of genes originating in the Z+E ancestor are GRAS
615 (Cheng et al. 2019), containing pro-orthologs of SCARECROW (SCR), SCARECROW-like, and
616 SHORTROOT transcription factors, which in land plants are essential to control cell division orientation
617 and tissue formation and they have also been associated with abiotic stress response. These genes allow
618 branching of cell filaments (i.e., formative cell divisions), which are occasionally found in multicellular
619 Zygnematophyceae.

620 GRAS homologs co-express with genes involved in cell division, cell cycle regulation, and cell wall
621 functions (cluster 147, 38 and 93). All three clusters contain genes associated with abiotic stress
622 responses, such as an ELIP homolog (OG 97 is expanded in *Zygnema*), beta-glucosidase (OG 85
623 expanded in the Z+E ancestor), calcium cation channel (DMI1/Pollux/Castor) and other calcium signaling
624 components, and LRR receptor-like protein kinases. Links to various phytohormone pathways exist via
625 the co-expression of a regulatory protein of ethylene receptor activity (TPR1) with an ABA signal
626 transducer (AIP2) (cluster 93), and ABA4, gibberellin with auxin-related gene ARF10 (cluster 38).
627 ARF10 often co-expresses with other TFs, some of which (e.g., CAMTA) have also been linked to biotic
628 and abiotic stress responses (Xiao et al., 2021). Cluster 38 showcases links between photosynthesis, ABA
629 signaling (ABA4), and the xanthophyll cycle (LUT2) that can mitigate photosynthetic stress.
630 Photooxidative protection is also suggested by the co-expression of MPH2 photosynthetic acclimation
631 factor and various redox proteins. The involvement of GRAS transcription factors in developmental and
632 environmental signaling networks speaks of the evolution of a complex network to coordinate growth and
633 stress since the common ancestor of Z+E.

634 In gene co-expression analysis, various gene modules reflect cell division and development. They
635 included homologs of phragmoplastin (cluster 87), kinesins motor proteins (e.g., clusters 52, 87), spindle
636 assembly proteins (cluster 52), RAB GTPases (cluster 10, 87), SNARE components (clusters 52, 87),
637 components of the cargo complex (clusters 10, 87), and cell division-related protein kinases. Cluster 52,
638 deployed during long day (14h) cycles that likely reflects steady-state growth, includes cyclin and cyclin-
639 dependent protein kinases involved in cell cycle regulation. Cell wall formation genes often co-express,
640 including cellulose or 1,4-beta-glucan synthases or transferases (cluster 10), beta-glucosidase (cluster 10;
641 OG 85 expanded in the Z+E ancestor), or fucosyltransferases (clusters 10, 52; OG 89 expanded in
642 Zygnematophyceae; see also **Figure 3C**). The co-expression of plastid division genes (e.g., components

643 of MinD, MinE, FtsZ1 and FtsZ2, ARC6 regulatory protein) speak of the coordinated division of plastids.
644 Further indicators of active cell division are genes related to DNA packaging and segregation such as
645 DNA topoisomerase, DNA helicase complex, chromatin remodeling factors, and components of the
646 condensin I and II complexes. Co-expression of protein phosphatases (PP2A; cluster 87) and
647 methyltransferases (clusters 10 and 52; OG 369 expanded in Zygnematophyceae) point to a functional
648 link between stress-responsive and cell division genes.
649

650 **Symbiotic genes are not co-expressed in Zygnematophyceae**

651 The symbiotic association with fungi was one of the key innovations that allowed plants to colonize land
652 (Rich et al. 2021). Phylogenetic analyses of land plant genes known for their symbiotic functions revealed
653 four of such genes in Zygnematophyceae (Delaux et al. 2015). It was thus proposed that these genes could
654 either form a pathway directly exapted for symbiosis in Embryophyta or that their genetic interactions
655 evolved along with the symbiotic habit. All four genes were found in *Zygnema*: DMI2/SYMRK pro-
656 ortholog (Zci_05951, DMI1/POLUX (Zci_12099), DMI3/CCaMK (Zci_01672) and IPD3/CYCLOPS
657 (Zci_13230). Co-expression analyses show that these genes likely belong to different modules (clusters
658 134, 78, 172, and 159, respectively). Given the diversity in putative functions of these additional genes, it
659 is not possible to propose a function for these clusters. The lack of co-expression between these genes in
660 Zygnematophyceae supports the hypothesis that the evolution of symbiosis in Embryophytes recruited
661 genes from diverse pathways rather than directly co-opting an existing pathway into a new function
662 (Delaux et al. 2012).
663

664 **Plant-microbe interactions and calcium signaling**

665 The first layer of plant responses to microbes hinges on their perception by receptors of extracellular or
666 intracellular signals (Wang et al. 2019); some of the receptors also act in the perception of mutualistic
667 partners (Plett and Martin, 2018). When encountering pathogens, depending on the signal and recognition
668 either pattern-triggered (PTI) or effector-triggered (ETI) immune responses are initiated. PTI and ETI
669 have been traditionally separated; yet they share many signaling pathways and downstream responses
670 (Cunha et al. 2006, Wang et al. 2019, Yuan et al. 2021). In PTI, molecules such as chitin or flagellin are
671 first recognized by pattern-recognition receptors (PRRs), whereas ETI is triggered by pathogen isolate-
672 specific effectors recognized by nucleotide-binding LRR receptors (NLRs). Typical immune responses
673 are callose deposition and ROS production. In *Zygnema*, we observed LRR domain-containing genes
674 often co-expressed with two callose synthase homologs (CAL5; Zci_11575, Zci_11576; cluster 117),
675 SOBER1/TIPSY1, which suppress ETI (Zci_13211; cluster 163), and a TOM2B homolog (Zci_03483;
676 cluster 107), which in *Arabidopsis thaliana* is associated with multiplication of tobamoviruses (Tsujimoto
677 et al. 2003). LRR proteins also co-express with immunity-associated receptor-like protein kinases
678 (clusters 107, 128, 163) and a ROP-activating protein RenGAP (Zci_12487; cluster 107). Cluster 163 is
679 mostly deployed under UV and other stresses and shows coexpression of LRR and SOBER1/TIPSY1
680 with a protein of the photosystem II assembly (LPA1; Zci_08073) and a plastidal protease (EGY;
681 Zci_01876). Overlap of some degree of defense and UV-B stress response is shown in land plants
682 (VandenBussche et al. 2018). In cluster 128, which is deployed most strongly under desiccation
683 conditions at 4°C, some LRR proteins have been involved in cell division (BAK1; Zci_06073 and SRF3;
684 Zci_03265). An interesting pattern is the frequent co-expression of LRR protein-encoding and calcium
685 signaling genes (Fig. 4A). For example, the calcium sensor and kinase (CPK; Zci_12352) in cluster 128
686 or CDPKs in cluster 117; cluster 117, showcasing the “plant perceptron” concept, also features PP2Cs

687 and LRR proteins (in fact, the most connected node is a LRR protein). Indeed, calcium signaling has
688 recently been proposed to link plant PTI and ETI (Jacob et al. 2021; Bjornson & Zipfel, 2022) but is also
689 important in mutualistic interactions (Plett and Martin, 2018). Overall, this highlights the
690 interconnectivity of the signaling cascades in *Zygnema*.

691

692 **Ancestry and diversity of *Zygnema* MADS-box genes**

693 In flowering plants, MADS-box genes control many developmental processes, from root to flower and
694 fruit development. Of special relevance is a lineage of Type II MADS-box genes termed MIKC-type
695 genes, which encode transcription factors with a characteristic domain-structure that includes a keratin-
696 like (K) domain that facilitates dimerization and enables tetramerization of these transcription factors,
697 yielding Floral Quartet-like Complexes (FQCs) (Puranik et al., 2014; Theißen et al., 2016). The increase
698 and diversification of these factors during land plants evolution is tightly associated with the
699 establishment of evolutionary novelties, especially in seed plants (Theißen et al., 2016).

700 We found one MADS-box gene each in the *Zygnema* genomes, which are closely related and do not
701 encode K domains (**Figure S16**). In transcriptomic data (One Thousand Plant Transcriptomes Initiative,
702 2019, Sayers et al., 2021), however, we found MADS-box genes encoding a K domain in other
703 Zygnematophyceae including a *Zygnema* species, which form a separate clade. This suggests the presence
704 of two MADS-box gene in the Zygnematophyceae ancestor: (i) an ancestral Type II that did not acquire
705 the K-domain yet and is thus very likely unable to form FQCs, and (ii) the MIKC-type, for which in one
706 case FQC formation has already experimentally been demonstrated in vitro (Rümpler et al., 2022). The
707 clade of the K-domain encoding genes was apparently lost in the *Zygnema* species sequenced here
708 (**Figure S16**). The ancestor of Zygnematophyceae thus had a higher basic diversity of Type II MADS-box
709 genes than embryophytes, but the number of MADS-box genes in *Zygnema* is similarly low than in other
710 streptophyte algae (Tanabe et al., 1995; Nishiyama et al., 2018), indicating that the boost of MADS-box
711 genes is a “synapomorphy” of land plants.

712

713 **Deep evolutionary roots of phytohormone biosynthesis and signaling pathways**

714 Phytohormone biosynthesis and signaling networks have deep evolutionary roots. Except for gibberellins
715 and jasmonic acid that likely originated in land plants (Bowman et al. 2019), all other phytohormone
716 pathways were present in green lineage ancestors. Looking at the number of phytohormone-associated
717 genes, land plants have more homologs than algae, as expected for their more complex signaling
718 pathways (Wang et al., 2015), and Zygnematophyceae are overall similar to other streptophyte algae
719 (**Figure 5**). We therefore explored how phytohormone-related genes are woven into *Zygnema*'s co-
720 expression networks. It must be noted that the identification of homologs known to be involved in
721 phytohormone biosynthesis and signaling in land plants cannot account for pathway variations that might
722 happen in algae.

723

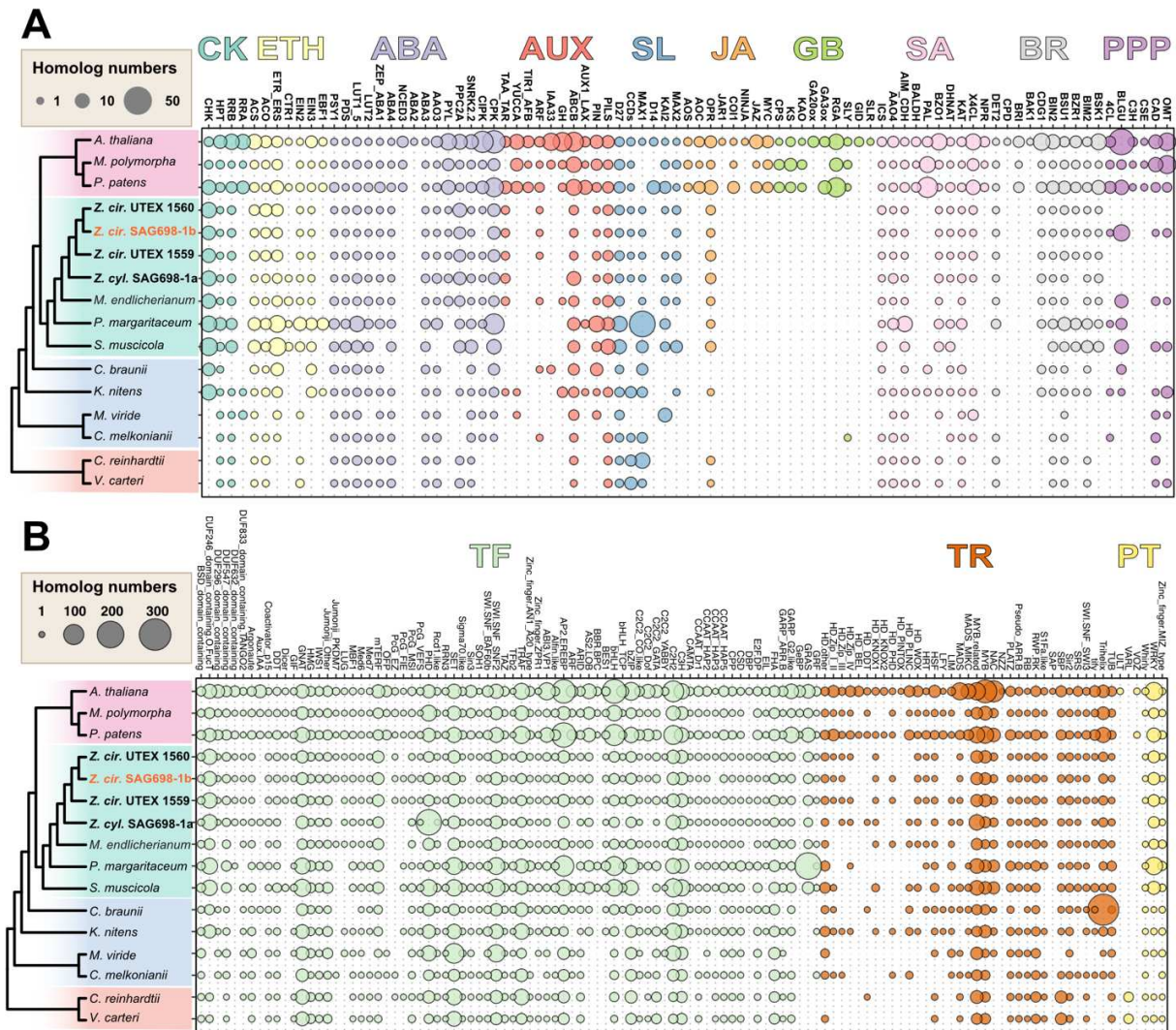


Figure 5. Phylogenetic distribution of (A) proteins involved in phytohormone biosynthesis, signaling, and phenylpropanoid biosynthesis, and (B) transcription factors. CK, cytokinin; ETH, ethylene; ABA, abscisic acid; AUX, auxin; SL, strigolactone; JA, jasmonic acid; GB, gibberellic acid; SA, salicylic acid; BR, brassinosteroids; PPP, phenylpropanoid; TF transcription factors; TR transcriptional regulators; PT putative transcription-associated proteins. For phytohormone-related proteins, homolog numbers were inferred from maximum likelihood gene family trees estimated from significant BLASTP hits (E-value < 1e-6) using *Arabidopsis* queries. Note that the high number of homologs found in *Penium margaritaceum* are likely due to the large genome of 3.6 Gb and >50K annotated proteins. Transcription factors were identified by TAPscan.

724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741

Cytokinin. Cytokinin is involved in plant growth and stress. In *Zygnema*, one out of four co-orthologs of the *A. thaliana* cytokinin receptor AHK2-4 contains the CHASE domain required for binding cytokinin (Zci_13126.1). Other CHASE-domain-containing histidine kinase (CHK) homologs result from more ancient duplications in streptophytes, where homologs previously identified *Spirogyra* and *Mougeotia* likely belong (de Vries et al. 2020). Two *Zygnema* CHKs were found to co-express with cell division and development genes such as cell cycle regulators, plate maturation factors, cell wall-relevant cellulose synthase, DNA replication factors and the only identified MADS-box gene (cluster 22, 76). Among response regulators, we found that the *Zygnema* strains lack Type-A *Arabidopsis* response regulators (RRA) and have only two Type-B *Arabidopsis* response regulators (RRB), and their involvement in

742 cytokinin signaling of Zygnematophyceae is unclear; in fact RRA showed low responsiveness to
743 exogenous cytokinin in *Spirogyra pratensis* (de Vries et al. 2020). Moreover, the paucity of RRA/RRB
744 genes in non-vascular plants supports the notion that this cytokinin-mediated induction of response
745 regulators might be a feature of seed plants (Brenner and Schmülling 2015). Yet, RRB homologs are co-
746 expressed with genes involved in cell division (cluster 151) and stress response (cluster 79).

747
748 **Ethylene.** In accordance with its deep evolutionary roots (Ju et al., 2015; Bowman et al., 2019), the
749 chassis for ethylene biosynthesis and signaling is present in Zygnematophyceae. EBF1, which binds EIN3
750 and is responsible for the last step in the signaling cascade, shows an almost land-plant specific
751 distribution, with the exceptions of EBF1, which is not found in *Zygnema* but homologs are found in
752 *Penium* and *Klebsormidium*. It is not yet known whether these algal homologs bind EIN3 in physiological
753 conditions. In land plants, ethylene triggers cell wall matrix modification, reduces chlorophyll
754 biosynthesis and photosynthesis, and activates abiotic stress responses. Exogenous ethylene treatment in
755 the zygnematophyte *Spirogyra* triggered similar transcriptomic responses and cell elongation (Van de
756 Poel et al., 2016) and *Spirogyra* genes can complement *Arabidopsis* ethylene signaling KO lines (Ju et al.,
757 2015). Several abiotic stress conditions have been shown to stimulate cell elongation in an ethylene-
758 dependent manner. Congruently, cluster 110, deployed under stress (high light, osmotic stress, UV, high
759 pH), shows ACS—a key regulator of ethylene biosynthesis—co-expressed with cell division and
760 circadian clock genes, genes related to photosystem II assembly, starch metabolism, and immunity. Two
761 homologs of ethylene receptor ETR1 co-express with genes involved in cell and plastid division, the PIN
762 auxin effector transporter, calcium signaling SIEL, a regulator of plasmodesmata intercellular trafficking
763 (cluster 119). Cluster 31, deployed largely under darker growing conditions (2h and 6h) and to a lesser
764 extent under high salt or high pH, shows co-expression of an ethylene-responsive factor with genes
765 involved in cell wall remodeling (transferases and beta-glucosidases), photosynthesis, and stress
766 responses (calcium signaling genes, subtilisin-like protease). These clusters illustrate well the effects of
767 ethylene in modifying the cell wall matrix, downregulating photosynthesis, and activating abiotic stress
768 responses as known from land plants.

769
770 **Abscisic acid.** Major aspects of the ABA signaling network are conserved across land plants (Cuming et
771 al., 2007; Umezawa et al., 2010; Eklund et al., 2018). The four new *Zygnema* genomes contain a complete
772 set of homologous genes to the ABA signaling cascade, including *Pyrabactin Resistance 1 /PYR1-like*
773 */Regulatory Component of ABA Receptor* (PYR/PYL/RCAR) receptors, as previously identified (de Vries
774 et al. 2018; Cheng et al., 2019). Functional data showed that Zygnematophyceae PYL regulates
775 downstream phosphatases in an ABA-independent manner (Sun et al., 2019). Among ABA biosynthetic
776 genes, two key *A. thaliana* enzymes lacked homologs outside of land plants (NCED3, Nine-cis-
777 Epoxy-carotenoid dioxygenase 3 and ABA2, SHORT-CHAIN DEHYDROGENASE/ REDUCTASE 1).
778 Yet, we detected about 1.01 ± 0.13 ng/g ABA in SAG 698-1b cultures by LC-MS (**Figure S13**), suggesting
779 that these reactions occur by alternative routes, perhaps *via* an ABA1-independent biosynthetic pathway
780 starting upstream of zeaxanthin as suggested by Jia et al. (2022). An interesting observation is the
781 expansion of clade A PP2Cs in *Zygnema* (5 homologs), akin to the 9 homologs found in *Arabidopsis*. The
782 expansion of PP2Cs was also detected by orthogroup expansion analyses (OG 548 and OG 830), but it
783 must be noted that the high PP2C numbers in *Arabidopsis* and *Zygnema* derive from independent
784 duplications.

785

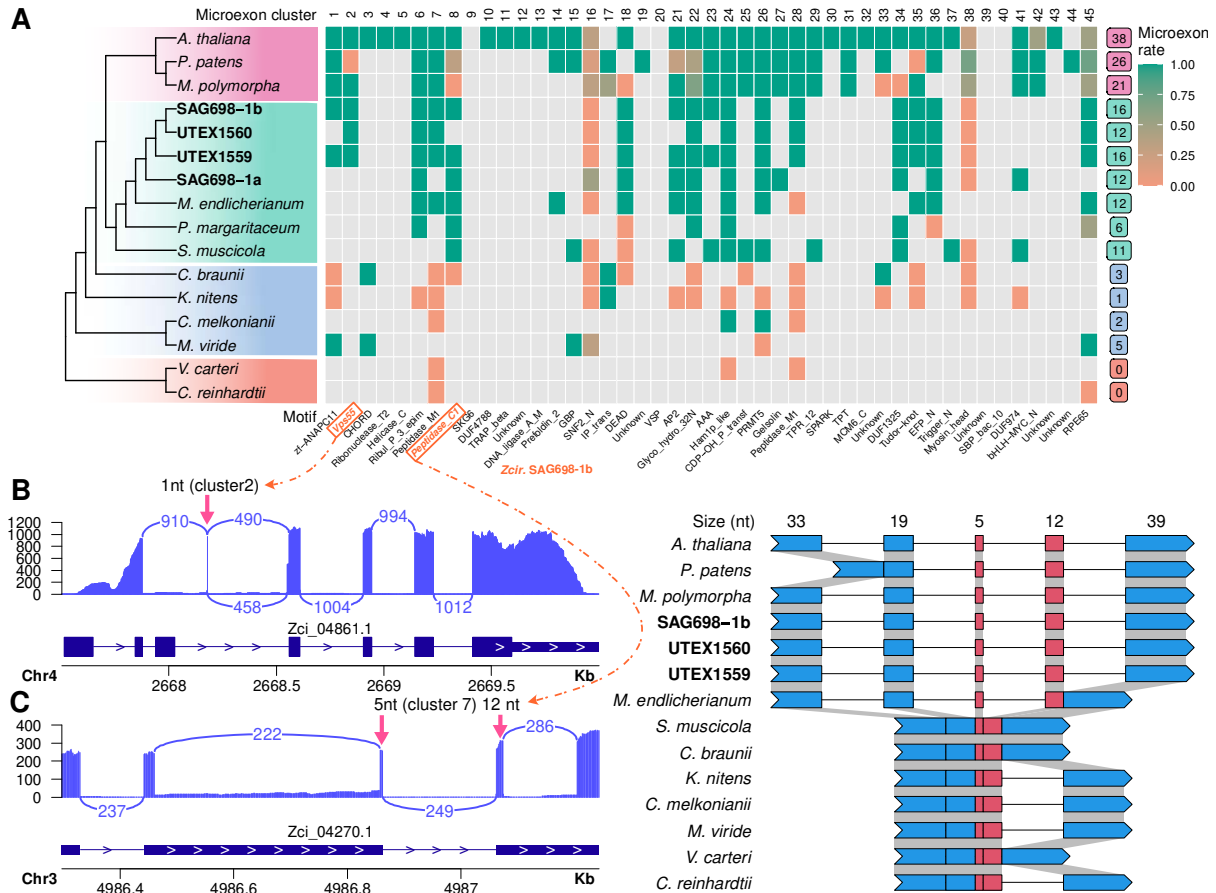
786 **Auxin.** Auxin is the major morphogenic phytohormone. Its polar distribution—largely based on the PIN
787 proteins (Adamowski and Friml, 2015)—leads to gradients along plant bodies, shaping various
788 developmental processes in all land plants (e.g., Friml et al., 2003; Weijers and Wagner, 2016). While all
789 components of the canonical auxin signaling pathway likely first came together in land plants (Flores-
790 Sandoval et al., 2018; Mutte et al., 2018; Martin-Arevalillo et al., 2019), PIN-mediated polar auxin
791 transport might have emerged earlier (Žabka et al., 2016); that said, not all streptophyte algal PINs
792 localize polarly (Skokan et al., 2019; Vosolsobě et al. 2020). While TAA homologs involved in auxin
793 biosynthesis are found in several streptophyte algae including *Zygnema*, no YUCCA homologs were
794 found in any Zygnematophyceae. In fact, embryophyte YUCCAs might have been acquired by HGT from
795 bacteria (Yue et al. 2014). Similarly, several signaling proteins (e.g., TIR, AUX, GH) are also absent from
796 *Zygnema* and distant homologs (never co-orthologs) are sometimes present in other streptophyte algae.
797 These patterns suggests that auxin signaling might be land plant specific.
798

799 **Strigolactones.** Strigolactone biosynthesis has been well established in land plants (Proust et al. 2011,
800 Delaux et al. 2012, Waters et al; 2017, Kodama et al. 2022). The streptophyte alga *Nitella* has been shown
801 to produce and respond to strigolactone although the actual biosynthetic pathway might differ from the
802 canonical one of embryophytes (Bowman et al. 2019, Nishiyama et al. 2018). Indeed, no clear orthologs
803 of carotenoid cleavage dioxygenases (CCD7 and CCD8; **Figure S10**), key biosynthetic enzymes, were
804 found in its transcriptome nor in the genome of the close relative *Chara braunii* (Nishiyama et al. 2018).
805 Strigolactones were also detected in some chlorophyte algal species for which no genomes data are
806 available (Smýkalová, I. et al. 2017). We find close homologs of strigolactone biosynthesis genes in all
807 *Zygnema* genomes (including CCD7 orthologs) as well as across the whole Chloroplastida. Yet, all
808 CCD7-8 homologs outside of embryophytes do not show conserved amino acid motifs proposed to be
809 important for substrate specificity (Messing, et al. 2010; **Figures S11, S12**). The angiosperm strigolactone
810 sensor D14 has likely evolved in the seed or vascular plant ancestors through neofunctionalization of
811 karrikin-sensing F-box proteins (Kodama et al. 2022) and thus no homologs are found in any algae. Our
812 results consolidate the hypothesis that strigolactone biosynthesis and signaling differ between
813 embryophytes and streptophyte algae.
814

815 **Microexons have evolved during plant terrestrialization**

816 Microexons are very short (1~15 basepairs (bp)) exons that can be evolutionarily conserved and crucial
817 for gene functions in plants (Yu et al., 2022). To study how microexons have evolved in streptophytes, we
818 predicted 45 microexon-tags in 16 plant genomes using *MEPmodeler* (<https://github.com/yuhuihui2011>).
819 Land plants typically have more than 20 of 45 microexon-tag clusters. In Zygnematophyceae genomes,
820 we found 10-20 microexon-tag clusters (only 6 clusters in *P. margaritaceum* probably due to the
821 fragmented genome assembly, **Table 1**), <5 in other streptophytes, and none in Chlorophyta (**Figure 6**).
822 Zygnematophyceae and land plants have the highest number of microexons. For example, a 1 bp
823 microexon of cluster 2 was found in Vps55 (Vacuolar protein sorting-associated protein 55, Zci_4861)
824 (**Figure 6B**), and two adjacent microexons (5 and 12 bp) of cluster 7 were found in a Peptidase M1 family
825 gene (Zci_04270) (**Figure 6C**), which are all supported by RNA-seq read mapping. Aligning the
826 orthologous genes of the Peptidase M1 family gene (Zci_04270) across multiple genomes, we found that
827 the two adjacent microexons are in the context of a 108 bp coding region spanning five exons in the
828 *Arabidopsis* gene (AT1G63770.5). The five-exon structure of this coding region is only conserved in land
829 plants and *Zygnema* (**Figure 6D**). In *M. endlicherianum* the last two exons (including the 112 bp) are

830 fused, while in earlier branching algae the five-exon structure exists as two or three exons with the two
 831 adjacent microexons (5 and 12 bp) of cluster 7 are always fused. It appears that during terrestrialization,
 832 at least for this Peptidase M1 family gene, there is a gradual intronization process that creates the higher
 833 abundance of microexons in land plants.
 834



835 **Figure 6:** Microexon prediction in 16 plant and algae genomes. (A) Heatmap of 45 conserved microexon-tags predicted by
 836 MEPmodeler with default parameters. Microexon rate is the rate of true microexons among all predicted results in the cluster,
 837 e.g., green cell indicates that 100% microexons with both two flanking introns are present, red indicates all microexon sequences
 838 are parts of large exons and none of them could be considered as microexons, and the others are between 0 and 1. A gray cell
 839 indicates missing data (a microexon-tag could not be found). Numbers on the right column indicate the predicted clusters
 840 containing at least one true microexon (see Yu et al., 2022 for more detail). (B) RNA-seq evidence of the 1 bp microexon in
 841 Cluster 2. (C) RNA-seq evidence of 1 bp microexon in Cluster 2 two adjacent microexons (5 and 12 bp) in Cluster 7. In B and C,
 842 the RNA-seq of condition p881sControl2 was used; RNA-seq read depth and gene annotation are shown; the number in each
 843 intron indicates the junction reads and the arrows point to microexons. (D) Exon-intron structures of microexon-tag Cluster 7 in
 844 14 plant genomes. The structure was predicted by relaxing the stringency in *M. viride* genome and by doing TBLASTX search in
 845 *S. muscicola* genome (all the three copies are intronless in this microexon-tag), respectively. The others are predicted with default
 846 parameters.
 847

849 Conclusion

850 One of the defining features of land plants is the plastic development of their multicellular bodies, ever
 851 adjusting to altered environmental conditions. We here generated chromosome-level genome assemblies
 852 for a representative of the filamentous algal sister lineage to land plants and performed exhaustive co-

853 expression network analyses. Our data underscore earlier notions of a deep evolutionary origin of
854 important plant signaling cascades for acclimation to environmental cues and suggest a deep conservation
855 of interconnections with regulation of growth. The plant perceptron connected environmental input with
856 development before embryophytes began their conquest of land.

857

858 **Acknowledgements**

859 This work was funded by the U.S. National Science Foundation (NSF) CAREER award (DBI-1652164),
860 the Nebraska Tobacco Settlement Biomedical Research Enhancement Funds as part of a start-up grant of
861 the University of Nebraska Lincoln, the Research & Artistry Award of Northern Illinois University, the
862 Joint Genome Institute Community Science Program (CSP), the United States Department of Agriculture
863 (USDA) award (58-8042-9-089), the National Institutes of Health (NIH) awards (R21AI171952) and
864 (R01GM140370) all to Y.Y., and by the German Research Foundation grant 440231723 (VR 132/4-1) to
865 J.d.V., TH417/12-1 to GT and FR, and 440540015 (BU 2301/6-1) to H.B. within the framework of the
866 Priority Programme “MAdLand – Molecular Adaptation to Land: Plant Evolution to Change” (SPP
867 2237), and grant 410739858 in the frame of the project CharMod to K.v.S., as well as RE 1697/16-1
868 (CharMod) and 18-1 (CharKeyS) to S.A.R. J.d.V. further thanks the European Research Council for
869 funding under the European Union’s Horizon 2020 research and innovation programme (Grant
870 Agreement No. 852725; ERC-StG “TerreStriAL”). The work was further supported by Austrian Science
871 Fund project P34181-B to A.H.; A.D., A.D.A., M.J.B., and J.M.S.Z. are grateful for being supported
872 through the International Max Planck Research School (IMPRS) for Genome Science; J.M.R.F.-J. and
873 T.P.R. gratefully acknowledge support by the Ph.D. program "Microbiology and Biochemistry" within
874 the framework of the “Göttingen Graduate Center for Neurosciences, Biophysics, and Molecular
875 Biosciences” (GGNB) at the University of Goettingen. The work (proposal:
876 10.46936/10.25585/60001088) conducted by the U.S. Department of Energy Joint Genome Institute
877 (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science
878 of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. P-M.D is
879 supported by the project Engineering Nitrogen Symbiosis for Africa (ENSA) currently funded through a
880 grant to the University of Cambridge by the Bill & Melinda Gates Foundation (OPP1172165) and the UK
881 Foreign, Commonwealth and Development Office as Engineering Nitrogen Symbiosis for Africa
882 (OPP1172165), by the "Laboratoires d'Excellence (LABEX)" TULIP (ANR-10-LABX-41)" and by the
883 European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation
884 program (Grant agreement No. 101001675).

885

886 **Authors’ contributions**

887 A.H., J.d.V., Y.Y. secured funding.

888 M.L., J.M.A., J.d.V., Y.Y. provided resources and materials.

889 A.H., J.d.V., Y.Y. provided supervision.

890 J.d.V., Y.Y. conceptualized the study.

891 J.d.V., K.B., Y.Y. performed project administration.

892 X.F., E.F., W.S.G., T.D., J.M.R.F.-J performed experimental work, including algal culturing, DNA/RNA
893 extraction, microscopy.

894 J.Z., X.F., and Y.Y. generated and annotated the draft genomes and transcriptomes.

895 I.I., S.d.V., J.d.V. analyzed phenylpropanoid metabolism-related genes.

896 R.D.H., I.V.G. coordinated genomes deposition/annotation in PhycoCosm.

897 L.B. performed and evaluated Illumina assemblies.
898 K.B., C.D. coordinated Illumina sequencing for UTEX 1559 and UTEX 1560.
899 L.G., F.R., and G.T. annotated and analyzed MADS-box genes
900 I.I., J.M.S.Z., T.P.R., A.D.A., A.A., A.M., P.M.-D. analyzed phytohormone-related genes.
901 J.B.A., N.K, A.M. performed ABA measurements.
902 H.B. analyzed data on cell division.
903 F.W.-L. analyzed photoreceptor genes.
904 K.v.S. helped in the initial phase of the project in strain purifications and mating experiments.
905 J.K. and P.-M.D. performed phylogenetic analyses of symbiotic genes
906 M.J.B., A.A. analyzed RNA-seq data.
907 X.F., J.Z., B.Z., T.L., O.N., I.I., J.d.V., Y.Y. analyzed data and generated data figures and tables.
908 J.Z., X.W., N.F.-P., S.A.R., Y.Y. conducted the WGD analysis.
909 S.A.R., R.P. performed TAPscan analysis.
910 S.A.R., F.H. performed contamination analyses.
911 N.R. and C.Pe. established the protocols and performed the chromosome stainings, staining
912 X.F., J.P.M. performed the organellar genome assembly and analysis.
913 X.F., J.Z., B.Z., J.H., Y.Y. performed the cell wall and HGT analysis.
914 H.Y., C.Z. performed the microexon analysis.
915 Z.A., M.M. built the *Zygnema* gene co-expression database.
916 X.F., I.I., J.Z., J.d.V., Y.Y. wrote the original draft.
917 All authors helped discussing the results and writing the paper.

918

919 **Declaration of Interests**

920 The authors declare no competing interests.

921

922 **METHODS**
 923
 924 **Resource table**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
DNasey Plant Pro Kit	QIAGEN, Hilden, Germany	Cat# 69204
DNasey PowerClean Cleanup Kit	QIAGEN, Hilden, Germany	Cat# 12877
Nanopore Ligation Sequencing Kit	Oxford Nanopore, Oxford, UK	Cat# SQK-LSK109
DNase I	Thermo Fisher, Waltham, MA, USA	
NEB mRNA stranded Library preparation kit	New England Biolabs, Beverly, MA, USA	
Trizol	Thermo Fisher, Waltham, MA, USA	
Deposited Data		
<i>Amborella trichopoda</i> genome	Amborella genome project, 2013	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Atrichopoda
<i>Arabidopsis thaliana</i> genome TAIR V10	TAIR	http://www.arabidopsis.org
<i>Azolla filiculoides</i> genome	Li et al., 2018	https://www.fernbase.org
<i>Bathycoccus prasinus</i> genome	Moreau et al., 2012	https://phycocosm.jgi.doe.gov/Batpra1/Batpra1.info.html
<i>Brachypodium distachyon</i>	The International <i>Brachypodium</i> Initiative 2010	https://phytozome-next.jgi.doe.gov/info/Bdistachyon_v3_1
<i>Chara braunii</i> S276 genome	Nishiyama et al., 2018	https://bioinformatics.psb.ugent.be/orcae/overview/Chbra
<i>Chlamydomonas reinhardtii</i> genome v5.5	Merchant et al., 2007; Blaby et al., 2014	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Creinhardtii
<i>Chlorokybus atmophyticus</i> genome	Wang et al., 2020	https://www.ncbi.nlm.nih.gov/assembly/GCA_009103225.1/
<i>Coccomyxa subellipsoidea</i> genome v2.0	Blanc et al., 2012	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_CsubellipsoideaC_169
<i>Coleochaete scutata</i> SAG 50.90 transcriptome	de Vries et al., 2018	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GFXZ000000000
<i>Klebsormidium nitens</i> genome v1.1	Hori et al., 2014	http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/
<i>Klebsormidium nitens</i> SAG2307 transcriptome	de Vries et al., 2018	https://www.ncbi.nlm.nih.gov/Traces/wgs/wgsviewer.cgi?val=GFXY000000000
<i>Marchantia polymorpha</i> genome v3.1	Bowman et al., 2017	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha
<i>Mesostigma viride</i> genome	Wang et al., 2020	https://www.ncbi.nlm.nih.gov/assembly/GCA_009103195.1/
<i>Mesotaenium endlicherianu</i> genome	Cheng et al., 2019	https://www.ncbi.nlm.nih.gov/assembly/GCA_009602735.1/
<i>Oryza sativa</i> Nipponbare genome v7.0	Kawahara et al., 2013	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Osativa
<i>Ostreococcus lucimarinus</i> genome v2.0	Palenik et al., 2007	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Olucimarinus
<i>Penium margaritaceum</i> genome	Jiao et al., 2020	http://bioinfo.bti.cornell.edu/cgi-bin/Penium/home.cgi
<i>Physcomitrium patens</i> genome v3.3	Lang et al., 2018	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ppatens

<i>Picea abies</i> genome	Nystedt et al., 2013	https://plantgenie.org/FTP?dir=Data%2FConGenIE%2FPicea_abies%2Fv1.0
<i>Salvinia cucullata</i> genome	Li et al., 2018	https://www.fernbase.org
<i>Selaginella moellendorffii</i> genome	Banks et al., 2011	https://phytozome-next.jgi.doe.gov/info/Smoellendorffii_v1_0
<i>Sphagnum fallax</i> v0.5 genome	Obtained from Phytozome with permission	https://phytozome-next.jgi.doe.gov/info/Sfallax_v0_5
<i>Spirogloea muscicola</i> genome	Cheng et al., 2019	https://www.ncbi.nlm.nih.gov/assembly/GCA_009602725.1
<i>Ulva mutabilis</i> genome	De Clerck et al., 2018	https://bioinformatics.psb.ugent.be/orcae/overview/Ulvmu
<i>Volvox carteri</i> genome v2.1	Prochnik et al., 2010	https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Vcarteri
<i>Zygnema cf. cylindricum</i> SAG 698-1a transcriptome	de Vries et al., 2018	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA399177
<i>Zygnema circumcarinatum</i> UTEX 1559 transcriptome	Fitzek et al., 2019	https://www.ncbi.nlm.nih.gov/sra/?term=SRX5449751
<i>Zygnema cf. cylindricum</i> SAG 698-1a genome	This study	https://phycocosm.jgi.doe.gov/SAG698-1a
<i>Zygnema circumcarinatum</i> SAG 698-1b genome	This study	https://phycocosm.jgi.doe.gov/SAG698-1b
<i>Zygnema circumcarinatum</i> SAG 1559 genome	This study	https://phycocosm.jgi.doe.gov/UTEX1559
<i>Zygnema circumcarinatum</i> SAG 1560 genome	This study	https://phycocosm.jgi.doe.gov/UTEX1560
<i>Zygnema circumcarinatum</i> SAG 698-1b co-expression networks	This study	https://zygnema.sbs.ntu.edu.sg/
Experimental Models: Organisms/Strains		
<i>Zygnema circumcarinatum</i> SAG 698-1b	Bohemia, ditch at meadow Poseltech (Poselský rybník) near Hirschberg (Doksy), 50.552702 / 14.669362 (800m), isolated by V. Czurda in 1929. Obtained from the Culture Collection of Algae at the University of Göttingen, Germany (SAG)	Maintained at the Culture Collection of Algae at the University of Göttingen, Germany (SAG) (https://sagdb.uni-goettingen.de/detailedList.php?str_number=698-1b)
<i>Zygnema circumcarinatum</i> UTEX 1559	Spontaneous mutant of IUCC 42 (UTEX B 42) with increased size and numbers of chloroplasts per cell, generated by H. Gauch, deposited in 1966. Obtained from the Culture Collection of Algae at University of Texas Austin	Maintained at the Culture Collection of Algae at the University of Texas at Austin (UTEX). (https://utex.org/products/utex-1559?variant=30991929671770)
<i>Zygnema circumcarinatum</i> UTEX 1560	Spontaneous mutant of IUCC 43 (UTEX B 43) with increased size and numbers of chloroplasts per cell, generated by H. Gauch, deposited in 1966. Obtained from the Culture Collection of Algae at University of Texas Austin	Maintained at the Culture Collection of Algae at the University of Texas at Austin (UTEX). (https://utex.org/products/utex-1560?variant=30991202058330)

<i>Zygnema cf. cylindricum</i> SAG 698-1a	Bohemia, ditch at meadow Poseltech (Poselský rybník) near Hirschberg (Doksy), 50.552702 / 14.669362 (800m), isolated by V. Czurda in 1929. Obtained from the Culture Collection of Algae at the University of Göttingen, Germany (SAG)	Maintained at the Culture Collection of Algae at the University of Göttingen, Germany (SAG) (https://sagdb.uni-goettingen.de/detailedList.php?str_number=698-1a)
Software and Algorithms		
ARWEN v1.2.3	Laslett and Canbäck, 2008	http://www.mybiosoftware.com/arwen-1-2-3-trna-detection-in-metazoan-mitochondrial-sequences.html
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
BUSCO v.5.2.2	Seppy et al. 2019	https://busco.ezlab.org
CAFE v5.0	Mendes et al., 2021	https://github.com/hahnlab/CAFE
canu	Koren et al., 2017	https://github.com/marbl/canu
dbCAN2	Zhang et al., 2018	https://github.com/linnabrown/run_dbcan
FASTQC	Babraham Institute	www.bioinformatics.babraham.ac.uk/projects/fastqc
FastTree 2.1	Price et al., 2009	http://www.microbesonline.org/fasttree/
Gblocks v0.91b	Castresana, 2000	http://molevol.cmima.csic.es/castresana/Gblocks_server.html
Geneious	Kearse et al., 2012	https://www.geneious.com/
Gseqseq	Tillich et al., 2017	https://chlorobox.mpimp-golm.mpg.de/gseqseq.html
GenomeScope	Vurture et al, 2017	http://qb.cshl.edu/genomescope/
HISAT2	Kim et al., 2019	https://daehwankimlab.github.io/hisat2/
hmmer	Finn et al., 2011	http://hmmer.org/
IQ-Tree v1.5.5	Nguyen et al., 2015	http://www.iqtree.org
iTOL v4.2.3	Letunic and Bork, 2021	https://itol.embl.de/
kakscalculator2	Wang et al., 2010	https://github.com/hahnlab/CAFE
kmergenie	Chikhi and Medvedev, 2014	http://kmergenie.bx.psu.edu/
MAFFT v7.3.10	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/source.html
MAKER-P v3.01.03	Campbell et al., 2014	http://www.yandell-lab.org/software/maker-p.html
MCSanX	Wang et al., 2012	https://github.com/wyp1125/MCSanX
MITE-tracker	Crescente et al., 2018	https://github.com/INTABiotechMJ/MITE-Tracker
minimap2	Li, 2018	https://github.com/lh3/minimap2
MUMmer v4.0.0	Marçais et al., 2018	https://github.com/mummer4/mummer
NOVOPlasty 3.8.2	Dierckxsens et al., 2017	https://github.com/ndierckx/NOVOPlasty
OGDRAW	Greiner et al., 2019	https://chlorobox.mpimp-golm.mpg.de/OGDraw.html
Orthofinder v2.5.2	Emms and Kelly, 2019	https://github.com/davidemms/OrthoFinder
PAML v4.10.0j	Yang, 2007	http://abacus.gene.ucl.ac.uk/software/paml.html
PASA v2.4.1	Haas et al., 2003	https://github.com/PASAPipeline/PASAPipeline
Pilon	Walker et al., 2014	https://github.com/broadinstitute/pilon
Phytools	Revell 2012	https://cran.r-project.org/web/packages/phytools/index.html
Platanus_allele	Kajitani et al., 2019	http://platanus.bio.titech.ac.jp
Posterior Mean Site	Wang et al. 2018	Implemented in IQ-Tree http://www.iqtree.org

Frequency Profiles		
Spades	Antipov et al., 2016	https://github.com/ablab/spades
Racon	Vaser et al., 2017	https://github.com/isovic/racon
RAxML v8	Stamatakis, 2014	https://cme.hits.org/exelixis/web/software/raxml/index.html
RepeatMasker 4.0.9	Bergman and Quesneville, 2007	http://www.repeatmasker.org/
RepeatModeler v2.0.1	Robert Hubley & Arian Smit	https://www.repeatmasker.org/RepeatModeler/
Re-routing method according to Yang 1995	Yang (1995)	n/a
Rfam 13.0	Kalvari et al., 2018	https://rfam.org/
StringTie	Pertea et al., 2016	https://ccb.jhu.edu/software/stringtie/
Transdecoder v.5.5.0	Brian J. Haas	https://github.com/TransDecoder/TransDecoder/releases
Trimal v1.4.rev15	Capella-Gutierrez et al. (2009)	http://trimal.cgenomics.org
Trimmomatic v0.36	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
Trinity v2.9.0	Grabherr et al., 2011	https://github.com/trinityrnaseq/trinityrnaseq
tRNAscan-SE v2.0	Lowe and Chan, 2016	http://lowelab.ucsc.edu/tRNAscan-SE/
WENGAN v0.2	Di Genova et al., 2021	https://github.com/adigenova/wengan
wtdbg2	Ruan and Li (2020)	https://github.com/ruanjue/wtdbg2

925

926 **Resource availability**

927 *Lead contact*

928 Further information and requests for resources and reagents should be directed to and will be fulfilled by
929 the lead contacts, Jan de Vries (devries.jan@uni-goettingen.de) and Yanbin Yin (yyin@unl.edu).

930

931 **Materials availability**

932 This study did not generate new unique reagents.

933

934 **Data and code availability**

935 The four *Zygnema* genomes, raw DNA reads of all four strains, and rRNA-depleted RNA-seq of
936 SAG 689-1b can be accessed through NCBI BioProject PRJNA917633. The raw DNA read data
937 of UTEX1559 and UTEX1560 sequenced by the Joint Genome Institute can be accessed through
938 BioProjects PRJNA566554 and PRJNA519006, respectively. RNA-seq data of UTEX1559 can
939 be accessed through BioProject PRJNA524229. Poly-A enriched RNA-seq data of SAG 689-1b
940 can be accessed through BioProject PRJNA890248 and the Sequence Read Archive (SRA) under
941 the accession SRR21891679 to SRR21891705. *Zygnema* genomes are also available through the
942 PhycoCosm portal (Grigoriev et al., 2021; <https://phycocosm.jgi.doe.gov/SAG698-1a>;

943 <https://phycocosm.jgi.doe.gov/SAG698-1b>; <https://phycocosm.jgi.doe.gov/UTEX1559>;
944 <https://phycocosm.jgi.doe.gov/UTEX1560>). No original code was used; all computational
945 analyses were performed with published tools and are cited in the Methods section.

946

947 **Organisms**

948 **Algal strains**

949 *Z. circumcarinatum* SAG 698-1b and *Z. cf. cylindricum* SAG 698-1a were obtained from the
950 Culture Collection of Algae at Göttingen University (SAG) (<https://sagdb.uni-goettingen.de>). *Z.*
951 *circumcarinatum* UTEX 1559 and UTEX 1560 were obtained from the UTEX Culture
952 Collection of Algae at UT-Austin (<https://utex.org/>). The algae were cultured with Bold's basal
953 media (BBM) or modified Bold's basal media (MBBM), supplemented with 0.02% L-arginine,
954 0.1% peptone and 0.5% sucrose (Feng et al., 2021). The filaments were grown for two or three
955 weeks on a rotary shaker platform (Fermentation Design, 125 rpm) in Plant Growth Chambers
956 (Convion PGR15) with conditions: 16:8 of light: dark cycle, 20°C, ~50 μmol of quanta $\text{m}^{-2} \cdot \text{s}^{-1}$
957 (Feng et al., 2021). Some cultures were also maintained on 1% agar or liquid MBBM in the same
958 conditions.

959

960 **Chromosome staining**

961 *Zygnema circumcarinatum* SAG 698-1b were collected at the end of the light phase and pre-
962 fixed in 2 mM 8-hydroxyquinoline (Roth, Karlsruhe, Germany) for 1h at room temperature (RT)
963 and 1h at 4°C and then fixed in Carnoy's fluid (3 parts 100% ethanol, 1 part 100% acetic acid)
964 for 12h at 4°C. Then, the totally bleached cells were stained in 1% acetocarmine (in 99% acetic
965 acid, Morphisto, Offenbach, Germany) by boiling for 5 min over open flame. Samples were then
966 transferred to microscope slides in 1% acetocarmine and examined at a Zeiss Axiovert 200 M
967 microscope (Carl Zeiss, Jena, Germany) and images captured with an Axiovision HRc camera
968 (Carl Zeiss, Jena, Germany). Stacked models of 100-150 images were rendered by Helicon
969 Focus software (HeliconSoft Ltd., Kharkiv, Ukraine) and further processed with ImageJ software
970 (Fiji, open source).

971

972 **Chromosome counting of *Z. cricumcarinatum* SAG 698-1b**

973 *Z.* strain SAG 698-1b was grown in liquid Bold's basal medium (BBM) in axenic culture under
974 controlled conditions (20°C, ~50 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$). Exposition to a 10:14 h light-dark
975 cycle led to synchronization of the cell cycle. The period where most cell divisions took place
976 was from the last hour of the light cycle to the first 5 h of the dark cycle, as previously reported
977 cell division occurs almost solely in the dark phase in *Zygnema* (Staker, 1971). After a two-week
978 incubation, algal filaments were harvested and pretreated for 1 h at RT, followed by 1 h at 4°C
979 in 2 mM 8-hydroxychinolin (Roth, Karlsruhe, Germany) in darkness, resulting in
980 depolymerization of microtubules and an increased condensation of chromosomes. Cells were
981 fixed in 1:3 glacial acetic acid - ethanol solution (Carnoy's fluid) for 12 h until all chlorophyll
982 was extracted. Fixed cells were stored in 70% ethanol at -20°C until staining, which was
983 performed by boiling in acetocarmine (99% acetic acid, 1% carmine; Morphisto, Offenbach,
984 Germany) for 5 min. To maximize the visualization of the chromosomes the filaments on the
985 prepared slides were slightly crushed, and stacks of 50 to 100 images per filaments with well
986 stained chromosomes were taken at a Zeiss Axiovert 200 M microscope (100x, 1.3 NA,
987 objective lens) with a Zeiss AxioCam HRm Rev.3 camera (Carl Zeiss, Jena, Germany). Stacked
988 models were rendered by the software Helicon Focus (HeliconSoft Ltd.). Final count of
989 chromosomes was done with ImageJ.

990

991 **DNA extraction**

992 Detailed protocols have been published elsewhere (Fitzek et al., 2019) (Orton et al., 2020) (Feng
993 et al., 2021). Briefly, algae were grown for two weeks and harvested using a vacuum filtration
994 with Whatman #2 paper (GE Healthcare 47 mm), washed with distilled water (three times),
995 frozen in liquid nitrogen and stored in -80°C. Frozen algae were lyophilized overnight and total
996 genomic DNA was extracted with DNeasy PowerPlant Pro Kit (Qiagen, Germany) using the
997 following workflow: lyophilized algae were first chopped with a spatula into fine powder and
998 mixed well with beads solution and RNase A. The mixture was homogenized on a vortex adapter
999 at maximum speed for 5 min. Then the DNeasy PowerPlant Pro Kit protocol was followed and

1000 the extracted DNA was further purified with DNeasy PowerClean CleanUp Kit (Qiagen,
1001 Germany).

1002 Before DNA extraction, to reduce chloroplast and mitochondria derived DNA (up to > 60%
1003 of total DNA), a modified nucleus isolation method was used (Zhang et al., 2012). Briefly, fresh
1004 algae tissues were grinded into fine powder in liquid nitrogen with precooled mortar and pestle.
1005 After that, the powder was transferred into a beaker containing nucleus isolation buffer. This
1006 mixture was homogenized well on ice, and then were vacuum filtered with two layers of
1007 miracloth (Thermo Fisher Scientific, USA). The remained nuclei were pelleted by centrifugation
1008 with speed of $\times 800$ g at 4 °C for 10 min, and extracted with DNeasy PowerPlant Pro Kit
1009 (Qiagen, Germany).

1010 After DNA extraction, quality and quantity of purified DNA was evaluated by using 1% agarose
1011 gel electrophoresis, NanoDrop 2000/2000c Spectrophotometers, and Qubit 3.0 Fluorometer
1012 (Thermo Fisher Scientific).

1013

1014 **Stress treatments and RNA-seq**

1015 We subjected *Z. circumcarinatum* SAG 698-1b to 19 growing and stress conditions, after which
1016 RNA-seq was obtained for the construction of a gene co-expression network. Stress and RNA-
1017 seq experiments were done in three batches. The first batch followed Pichrtová et al. (2014) and
1018 de Vries et al. (2018) with modifications. Three-week algae were sub-cultured in 12 flasks of
1019 liquid BBM with 0.02% L-arginine and grown for two weeks under standard conditions: 16/8 of
1020 light/dark cycle at 20°C and $\sim 50 \mu\text{mol}$ of quanta $\text{m}^{-2} \cdot \text{s}^{-1}$. Then the algae were treated for 24 h
1021 under four conditions: (i) 20°C in liquid medium (standard control); (ii) 4°C in liquid medium,
1022 (iii) desiccation at 20°C, and (iv) desiccation at 4°C. Four treatments each with three replicates
1023 were performed. For desiccation treatments, algae were harvested using a vacuum filtration with
1024 Glass Microfiber Filter paper (GE Healthcare, 47 mm) and 20 μl of MBBM was added on the
1025 filter paper. Papers with algae were then transferred onto a glass desiccator containing saturated
1026 KCl solution (Pichrtová et al., 2014) and the desiccator was sealed with petroleum jelly and
1027 placed in the growth chamber under standard culture conditions. Cultures grown in liquid
1028 conditions were harvested using a vacuum filtration with Whatman #2 paper (GE Healthcare, 47

1029 mm). After 24 h of treatment, the twelve samples were transferred into 1.5 ml Eppendorf tubes
1030 and immediately frozen in liquid nitrogen and stored in -80°C . For the second batch (6 diurnal
1031 experiments), the algae were grown with the same control conditions as the above mentioned
1032 (16/8 of light/dark cycle, 20°C , $\sim 50 \mu\text{mol}$ of quanta $\text{m}^{-2} \cdot \text{s}^{-1}$), and samples were collected every
1033 4 hours: (v) diurnal dark 2h, (vi) diurnal dark 6h, (vii) diurnal light 2h, (viii) diurnal light 6h, (ix)
1034 diurnal light 10h, (x) diurnal light 14h. For the third batch (9 stress experiments): SAG 698-1b
1035 was pre-cultivated at 20°C , 16:8 hrs light:dark cycle at $90 \mu\text{mol}$ photons/ m^2s on a cellophane
1036 discs (folia Bringmann, Germany) for 8 days. For certain treatments (NaCl, Mannitol,
1037 CadmiumCl) the culture was transferred to a new petri dish where the medium was
1038 supplemented with the substances mentioned above. Algae were then subjected to: (xi) $150 \mu\text{M}$
1039 NaCl (Roth, Germany) for 24h, (xii) 300mM Mannitol (Roth, Germany) for 24h, (xiii) $250 \mu\text{M}$
1040 CadmiumCl (Riedel-de Haën AG, Germany) for 24 h, (xiv) dark treatment for 24h, (xv) high
1041 light (HL) treatment at $900 \mu\text{mol}$ photons/ m^2s for 1h, (xvi) UV-A at 385nm , $1400 \mu\text{W}/\text{cm}^2$ for
1042 1h, (xvii) high light at 4°C (HL4) at $900 \mu\text{mol}$ photons/ m^2s for 1h, (xviii) $\text{pH}=9$ for 24h, and
1043 (xix) a corresponding control growth at 20°C on a plate.

1044

1045 **RNA extraction**

1046 Frozen algae were lyophilized overnight for RNA extraction with a modified CTAB method
1047 described in Chang et al. (1993) and Bekesiova et al. (1999). Specifically, the tissue was
1048 chopped with spatula into fine powder, and then $500 \mu\text{l}$ of CTAB buffer (2% CTAB, 100mM
1049 Tris-HCl pH 8.0, 25mM EDTA, 2M NaCl, 2% Polyvinylpyrrolidone, 1% β -mercaptoethanol)
1050 was added and mixed well. The mixture was incubated in heating block at 65°C for 15 min.
1051 After the tubes cooled down, the solution was extracted with Chlorophorm: Isoamyl alcohol 24:1
1052 twice. The supernatant was precipitated with 0.3 volume of 10M LiCl that was incubated in $-$
1053 20°C for 30 min. The pellet was washed with 75% ethanol twice and vacuum dried for 15 min.
1054 RNA was resuspended in $50 \mu\text{l}$ of 0.1% DEPC (diethylpyrocarbonate) water. RNA samples were
1055 treated with RNase-Free DNase I (Promega) at 37°C for 30 min to remove any DNA residue.
1056 Quality and quantity of purified RNA was evaluated by using 1% agarose gel electrophoresis,

1057 NanoDrop 2000/2000c Spectrophotometers, Qubit 3.0 Fluorometer (Thermo Fisher Scientific)
1058 and RNA Integrity Number (RIN) (Agilent).

1059

1060 **Library preparation and sequencing**

1061 DNA samples were sequenced at Roy J. Carver Biotechnology Center at University of Illinois at
1062 Urbana-Champaign, using Oxford Nanopore and Illumina technologies (Table S1A). Oxford
1063 Nanopore DNA libraries were prepared with 1D library kit SQK-LSK109 and sequenced with
1064 SpotON R9.4.1 FLO-MIN106 flowcells for 48h on a GridIONx5 sequencer. Base calling was
1065 performed with Guppy v1.5 (<https://community.nanoporetech.com>). Illumina shotgun genomic
1066 libraries were prepared with the Hyper Library construction kit (Kapa Biosystems, Roche).
1067 Libraries had an average fragment size of 450 bp, from 250 to 500 bp, and sequenced with 2x250
1068 bp pair-end on HiSeq 2500. Additional DNA samples were sequenced at the Genome Research
1069 Core in University of Illinois at Chicago and JGI. The Illumina shotgun genomic libraries were
1070 prepared with Nextera DNA Flex Library Prep Kit. The libraries had an average fragment size of
1071 403bp and sequenced with 2x150 bp pair-end on HiSeq 4000 (Table S1A). RNA samples were
1072 sequenced at the Genome Research Core in University of Illinois at Chicago. The libraries were
1073 prepared by rRNA depletion with Illumina Stranded Total RNA kit plus Ribo-Zero Plant
1074 ([https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-stranded-](https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-stranded-total-rna-plant.html)
1075 [total-rna-plant.html](https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-stranded-total-rna-plant.html)), and 2x150 bp pair-end sequencing was performed on HiSeq 4000. RNA
1076 from the third batch of stress experiments were sequenced at the NGS-Integrative Genomics
1077 Core Unit of the University Medical Center Göttingen, Germany. Stranded mRNA libraries were
1078 prepared with the Illumina stranded mRNA kit and paired-end sequencing of 2x150 bp reads was
1079 carried out on an Illumina HiSeq4000 platform.

1080 RNA-seq data for SAG 698-1a and UTEX 1559 have been previously published (Table
1081 S1A).

1082

1083 **Transcriptome assembly**

1084 Raw RNA-seq reads (Table S1A) were quality checked with FastQC v.0.11.9
1085 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) (Andrews, 2010), trimmed with

1086 TrimGalore (<https://github.com/FelixKrueger/TrimGalore>), and were inspected again with
1087 FastQC. All reads were combined, and *de novo* assembled with Trinity version 2.9.0 (Grabherr et
1088 al., 2011; Haas et al., 2013).

1089

1090 **K-mer frequency analysis**

1091 The trimmed DNA Illumina reads were filtered out with BLASTP using plastomes and
1092 mitogenomes from *Zygnema* as references. Remaining (putatively nuclear) were used to predict
1093 the best k-mer size by kmergenie (<http://kmergenie.bx.psu.edu/>) (Chikhi and Medvedev, 2014).
1094 The histogram of the best k-mer was then uploaded to GenomeScope for viewing the genome
1095 plot (<http://qb.cshl.edu/genomescope/>) (Vurture et al., 2017) (Table S1B and Figure S2).

1096

1097 **Genome assembly and scaffolding**

1098 To assemble the genome of SAG 698-1b, a total of 5.4 Gb (82x) of Oxford Nanopore nuclei
1099 DNA reads were assembled with wtdbg (Ruan and Li, 2020) (<https://github.com/ruanjue/wtdbg>).
1100 Assembled contigs were polished by Racon (Vaser et al., 2017) and three iterations of pilon
1101 (Walker et al., 2014) with Illumina paired-end reads. The polished genome was scaffolded by
1102 Dovetail Genomics HiRise software with Hi-C sequencing data (<https://dovetailgenomics.com/>).
1103 Genome contamination was examined by BLASTX against NCBI's NR database and
1104 contaminated scaffolds were removed.

1105 To assemble the UTEX 1559 genome, an initial assembly was done with SPAdes (Antipov
1106 et al., 2016) using Illumina paired-end reads (2x150), three mate-pair libraries (insert size: 3-5
1107 kb; 5-7 kb and 8-10 kb) and Oxford Nanopore reads (Table S1A). Assembled contigs were
1108 further scaffolded by two rounds of Platanus (Kajitani et al., 2019) with Illumina paired-end
1109 reads (2x 250), three mate-pair libraries (insert size 3-5 kb; 5-7 kb and 8-10 kb) and Oxford
1110 Nanopore reads. For the UTEX 1560 genome, Illumina paired-end (2x150 bp) and PacBio HiFi
1111 reads were used for assembly with SPAdes and further scaffolded with Platanus. Scaffolds with
1112 contaminations were identified by BLASTX against NR and removed. The genomes of
1113 UTEX1559 and UTEX1560 were scaffolded by Dovetail Genomics HiRise software with Hi-C
1114 sequencing data from SAG 698-1b.

1115 The genome of SAG 698-1a was sequenced with PacBio HiFi long reads (40Gb), Nanopore
1116 long reads (4Gb), and Illumina short reads (>100Gb). The k-mer analysis using Illumina reads
1117 revealed two peaks in the k-mer distribution, suggesting that SAG 698-1a exists as a diploid
1118 organism with an estimated heterozygosity rate of 2.22% (Figure S2). All Illumina short reads
1119 and the Nanopore reads were first assembled into contigs using SPAdes. Then, WENGAN was
1120 used to assemble HiFi long reads and Illumina paired-end reads (2x150bp) using the SPAdes
1121 contigs as the reference. Lastly, the resulting WENGAN contigs were scaffolded and gaps were
1122 closed with Platanus-allee using all the Nanopore, HiFi, and Illumina reads to derive a consensus
1123 pseudo-haploid genome.

1124 To evaluate the quality of assembled genomes (Table S1D), raw RNA-seq reads, Oxford
1125 Nanopore and Illumina DNA reads were mapped to the assembly with hisat2 (Kim et al., 2019),
1126 minimap2 (Li, 2018), and bowtie2 (Langmead and Salzberg, 2012), respectively. To assess
1127 genome completeness, a BUSCO (Seppey et al., 2019) analysis was performed with the
1128 ‘Eukaryota odb10’ and ‘Viridiplantae odb10’ reference sets.

1129

1130 **Repeat annotation**

1131 Repetitive DNA was annotated using the homology strategy with repeat libraries generated with
1132 RepeatModeler. RepeatModeler integrates RepeatScout, RECON, LTRharvest and
1133 LTR_retriever tools (version 2.0.1; <http://www.repeatmasker.org/RepeatModeler/>) (Flynn et al.,
1134 2020). The MITE (Miniature inverted-repeat transposable elements) library was identified with
1135 MITE-tracker (Crescente et al., 2018) software. These two identified libraries were combined
1136 and incorporated into RepeatMasker (version 4.0.9; <http://www.repeatmasker.org/>) for repeat
1137 annotation.

1138

1139 **Genome annotation**

1140 In all four genomes, protein coding genes were predicted by the MAKER-P pipeline (Campbell
1141 et al., 2014) which integrates multiple gene prediction resources, including *ab initio* prediction,
1142 protein homology-based gene prediction and transcripts-based evidence. First, repetitive
1143 elements were masked by RepeatMasker with a custom repeat library generated by

1144 RepeatModeler. Rfam with infernal and tRNA-Scan2 were used to analyze non-coding RNA and
1145 tRNA. For the transcripts evidence, total of 103,967 transcripts were assembled by Trinity
1146 (reference-free) and StringTie (reference-based) with RNA-seq reads. Transcriptome assembly
1147 were used for generating a complete protein-coding gene models using PASA. Proteins from *M.*
1148 *endlicherianum*, *S. muscicola* and *A. thaliana* (TAIR10) were used as homology-based evidence.
1149 Then, the resulting protein-coding gene models from the first iteration of MAKER-P pipeline
1150 were used as training data set for SNAP and Augustus models, which were fed into MAKER for
1151 the second iteration of annotation. After three rounds of gene prediction, MAKER-P combined
1152 all the protein-coding genes as the final annotated gene sets.

1153

1154 **Plastome and mitogenome assembly and annotation**

1155 NOVOPlasty 3.8.2 (<https://github.com/ndierckx/NOVOPlasty>) (Dierckxsens et al., 2017) was
1156 used to assemble plastomes. The contiguity of assembled plastomes was examined in Geneious
1157 software (<https://www.geneious.com/>) (Kearse et al., 2012) with read mapping. For SAG 698-1b
1158 mitogenome assembly, Oxford Nanopore reads were assembled with Canu (Koren et al., 2017),
1159 where one long mitogenome contig of 238,378 bp was assembled. This contig was circularized
1160 and polished with three rounds of pilon (Walker et al., 2014), that was further corrected with
1161 Illumina raw reads and compared with mitogenome of UTEX 1559 (MT040698; Orton et al.,
1162 2020) in Geneious. For SAG 698-1a, PacBio HiFi reads were used for the assembly of its
1163 mitogenome.

1164 Plastome and mitogenome annotation was performed with GeSeq (Tillich et al., 2017)
1165 (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>). For plastome annotation, BLAT search and
1166 HMMER profile search (Embryophyta chloroplast) were used for coding sequence, rRNA and
1167 tRNA prediction; ARAGORN v1.2.38, ARWEN v1.2.3 and tRNAscan-SE v2.0.5 were used for
1168 tRNA annotation. For mitogenome annotation, Viridiplantae was used for BLAT Reference
1169 Sequences. The annotated gff files were uploaded for drawing circular organelle genome maps
1170 on OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) (Greiner et al., 2019).

1171

1172 **Comparison of *Z. circumcarinatum* genomes (SAG 698-1b, UTEX 1559, UTEX 1560)**

1173 Two approaches were used to compare the three genomes (Figure S8). The first approach was
1174 based on the whole genome alignment (WGA) by using MUMMER. The parameters "--
1175 maxmatch -c 90 -l 40" were set to align the three genomes and then "-i 90 -l 1000" were set to
1176 filter out the smaller fragments. The second approach focused on the gene content comparisons.
1177 Orthofinder was used to obtain ortholog groups (orthogroups) from genomes' annotated proteins.
1178 Orthofinder results led to a Venn diagram with unique genes (orthogroups with genes from only
1179 1 genome), cloud genes (orthogroups with genes from only 2 genomes), and core genes
1180 (orthogroups with genes from all 3 genomes), which collectively form the pan-genome.
1181 Orthofinder could have failed to detect homology between very rapidly evolved orthologous
1182 genes, which leads to an under-estimation of core genes. Also, gene prediction may have missed
1183 genes in one genome but found them in other genomes. To address these issues, the raw DNA
1184 reads of each genome were mapped to the unique genes and cloud genes using BWA. This step
1185 was able to push more unique genes and cloud genes to core genes or push some unique genes to
1186 cloud genes. The following criteria were used to determine if an orthogroup needed to be re-
1187 assigned: (i) the number of reads and coverage calculated by bedtools are >10 and >0.8 for a
1188 gene, respectively, and (ii) >60% coverage of genes in the orthogroup find sequencing reads
1189 from the other genomes. After this step the final Venn diagram was made (Figure S8F), showing
1190 the counts of the final core genes, cloud genes, and unique genes.

1191

1192 **Whole genome duplication (WGD) analysis**

1193 To identify possible WGDs, Ks and 4dTV values were calculated for each genome. First, all
1194 paralog pairs were identified using RBBH (Reciprocal Best BLAST Hit) method using protein
1195 sequences (E-value < 1e-6), following the method described by Bowman et al. (2017). RBBH
1196 paralog pairs were aligned with MAFFT (Katoh and Standley, 2013) and the corresponding
1197 nucleotide alignments were generated. Using RBBH alignments of paralog pairs,
1198 KaKs_Calculator2.0 (Wang et al., 2010b) with the YN model and the
1199 calculate_4DTV_correction.pl script were run to calculate Ks and 4dTV values for each
1200 alignment, respectively. Ks = 0 and 4dTV = 0 values were filtered. The Ks and 4dTV distributions
1201 were fitted with a gaussian kernel density model using the seaborn package. For the SAG 698-1b

1202 chromosome-level genome, MCscan (Wang et al., 2012) was run to identify syntenic block
1203 regions with default parameters.

1204

1205 **Species phylogeny and divergence time analysis**

1206 Sixteen representative genomes were selected, including two chlorophytes (*Volvox carteri*,
1207 *Chlamydomonas reinhardtii*), seven Zygnematophyceae (*Zygnema circumcarinatum* SAG 698-
1208 1b, UTEX 1559, UTEX 1560, *Z. cf. cylindricum* SAG 698-1a, *Mesotaenium endlicherianum*,
1209 *Penium margaritaceum*, *Spirogloea muscicola*), four additional streptophyte algae (*Chara*
1210 *braunii*, *Klebsormidium nitens*, *Chlorokybus melkonianii*, *Mesostigma viride*), two bryophytes
1211 (*Marchantia polymorpha*, *Physcomitrium patens*) and a vascular plant (*Arabidopsis thaliana*).
1212 Orthogroups were generated by OrthoFinder version 2.5.2 (Emms and Kelly, 2019) and 493
1213 low-copy orthogroups containing ≤ 3 gene copies per genome were aligned with MAFFT v7.310
1214 (Kato and Standley, 2013). Gene alignments were concatenated and gaps were removed by
1215 Gblocks version 0.91b (Castresana, 2000). Phylogenetic tree was built using RAxML v.8
1216 (Stamatakis, 2014) with the “-f a” method and the PROTGAMMAJTT model, and support with
1217 100 pseudoreplicates of non-parametric bootstrap. The tree was rooted on Chlorophyta.

1218 Using the above methodology, additional phylogenetic analyses were performed with (i) the
1219 four *Zygnema* strains and (ii) the seven Zygnematophyceae genomes, in order to obtain a higher
1220 number of single-copy loci, 5,042 and 204, respectively (**Figure S7**).

1221 Divergence time estimation was carried out with MCMCTree implemented in the PAML
1222 package version 4.10.0j (Yang, 2007). The 493 low-copy orthogroup protein sequence alignment
1223 was converted to the corresponding nucleotide alignment for MCMCTree, in which 10 MCMC
1224 (Markov Chain Monte Carlo) chains were run, each for 1,000,000 generations (Table S1F).
1225 Three calibration were set in the reference tree according to (Morris et al., 2018) on the nodes
1226 Viridiplantae (972.4~669.9 Ma), Streptophyta (890.9~629.1 Ma) and Embrophyta (514.8~473.5
1227 Ma).

1228

1229 **Gene family phylogenetic analysis**

1230 CAZyme families were identified with dbCAN2 (Zhang et al., 2018) with default parameters (E-
1231 value $< 1e-10$ and coverage > 0.35). Whenever needed, dbCAN2 was rerun by using more
1232 relaxed parameters. The experimentally characterized cell wall enzymes were manually curated
1233 from the literature (Data S1 and Table S1L). Reference genes were included into the phylogenies
1234 to infer the presence of orthologs across the 16 genomes and guide the split of large families into
1235 subfamilies. Phylogenetic trees were built by using FastTree (Price et al., 2009) initially, and for
1236 some selected families, RAxML (Price et al., 2009) and IQ-Tree (Nguyen et al., 2015) were used
1237 to rebuild phylogenies to verify topologies.

1238

1239 **Orthogrup expansion and contraction analysis**

1240 We inferred expanded and contracted gene families with CAFE v.5 using orthogroups inferred
1241 with Orthofinder v.2.4.0 and the previously inferred time-calibrated species phylogeny. CAFE
1242 v.5 was run with default settings (base) using the inferred orthogroups and a calibrated species
1243 phylogeny. Two independent runs arrived to the same final likelihood and lambda values. The
1244 first eight orthogroups (OG0-7) were excluded from the analysis due to too drastic size changes
1245 between branches that hampered likelihood calculation; excluded orthogroups were mostly
1246 exclusive to a single *Zygnema* or *Chara* genome and likely represented transposable elements, as
1247 judged by results of BLASTP against NR.

1248

1249 **Phytohormones**

1250 Proteins involved in phytohormone biosynthesis and signaling were identified by BLASTP
1251 against annotated proteomes (e-value $<1e-6$) using *Arabidopsis* genes as queries. For genes with
1252 ubiquitous domains (e.g. CIPK, CPK3, SNRK2, CDG1, BAK1), hits were filtered by requiring
1253 BLASTP coverage $\geq 50\%$ of the query. Significant hits were then aligned (MAFFT auto) and
1254 maximum likelihood gene trees were inferred in IQ-Tree using best-fit models and 1000
1255 replicates of SH-like aLRT branch support ('-m TEST -msub nuclear -alrt 1000'). The final sets
1256 of homologs were identified by visually inspecting gene trees and identifying the most
1257 taxonomically diverse clade (with high support of SH-aLRT >0.85) that included the
1258 characterized *Arabidopsis* proteins. Bubble plot was generated with ggplot2 in R.

1259

1260 **Screening for symbiotic genes and phylogenetic analysis**

1261 Symbiotic genes were screened against a database of 211 plant species across Viridiplantae
1262 lineage (see Table) using proteins of the model plant *Medicago truncatula* as queries in BLASTP
1263 v2.11.0+ (Camacho et al., 2009) searches with default parameters and an e-value < 1e-10. Initial
1264 alignments of all identified homologs was performed using the DECIPHER package (Wright,
1265 2015) in R v4.1.2 (R Core Team). Positions with >60% gaps were removed with trimAl v1.4
1266 (Capella-Gutiérrez et al., 2009) and a phylogenetic analysis was performed with FastTree
1267 v2.1.11 (Price et al., 2009). Clades corresponding to *M. truncatula* orthologs queries were
1268 extracted and a second phylogeny was performed. Proteins were aligned using MUSCLE
1269 v3.8.1551 (Edgar, 2004) with default parameters and alignments cleaned as described above.
1270 Tree reconstruction was performed using IQ-TREE v2.1.2 (Minh et al., 2020) based on BIC-
1271 selected model determined by ModelFinder (Kalyaanamoorthy et al., 2017). Branch supports
1272 was estimated with 10,000 replicates each of both SH-aLRT (Guindon et al., 2010) and UltraFast
1273 Bootstraps (Hoang et al., 2017). Trees were visualized and annotated with iTOL v6 (Letunic and
1274 Bork, 2021). For the GRAS family, a subset of 42 species representing the main lineages of
1275 Viridiplantae was selected (see Table) and GRAS putative proteins screened using the
1276 HMMSEARCH program with default parameters and the PFAM domain PF03514 from
1277 HMMER3.3 (Johnson et al., 2010) package. Phylogenetic analysis was then conducted as
1278 described above.

1279

1280 **Screening for CCD homologs and phylogenetic analysis**

1281 Annotated proteins from 21 land plant genomes (*Amborella trichopoda*, *Anthoceros agrestis*,
1282 *Anthoceros punctatus*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Azolla filiculoides*,
1283 *Brachypodium distachyon*, *Brassica rapa*, *Lotus japonicus*, *Marchantia polymorpha*, *Medicago*
1284 *truncatula*, *Oryza sativa*, *Physcomitrium patens*, *Picea abies*, *Pisum sativum*, *Salvinia cucullata*,
1285 *Selaginella moellendorffii*, *Sphagnum fallax*, *Spinacia oleracea*, *Gnetum montanum*, *Crocus*
1286 *sativus*), 7 streptophyte algal genomes (*Spirogloea muscicola*, *Penium margaritaceum*,
1287 *Mesotaenium 'endlicherianum'*, *Mesostigma viride*, *Klebsormidium nitens*, *Chlorokybus*

1288 *melkonianii*, *Chara braunii*, *Zygnema circumcarinatum*), 6 chlorophyte genomes (*Ulva*
1289 *mutabilis*, *Ostreococcus lucimarinus*, *Micromonas pusilla*, *Micromonas sp.*, *Chlamydomonas*
1290 *reinhardtii*, *Coccomyxa subellipsoidea*, *Chlorella_variabilis*), 5 cyanobacterial genomes
1291 (*Trichormus azollae*, *Oscillatoria acuminata*, *Nostoc punctiforme*, *Gloeomargarita lithophora*,
1292 *Fischerella thermalis*), as well as the transcriptome of *Coleochaete scutata* (de Vries et al.,
1293 2018). The representative *A. thaliana* protein was used as query for BLASTP searches against
1294 the above annotated proteins (E-value < 1e-5). Homologs were aligned with MAFFT v7.453 L-
1295 INS-I approach (Katoh and Standley, 2013) and maximum likelihood phylogenies computed
1296 with IQ-Tree v.1.5.5 (Nguyen et al., 2015), with 100 bootstrap replicates and BIC-selected
1297 model (WAG+R9) with ModelFinder (Kalyaanamoorthy et al., 2017). Functional residue
1298 analyses were done based on published structural analysis (Messing et al., 2010), and the
1299 alignments were plotted with ETE3 (Huerta-Cepas et al., 2016).

1300

1301 **Phylogeny of MADS-box genes**

1302 MADS-domain proteins were identified by Hidden Markov Model (HMM) searches (Eddy,
1303 1998) on annotated protein collections. MADS-domain sequences of land plants and
1304 opisthokonts were taken from previous publications (Marchant et al., 2022; Gramzow et al.,
1305 2010). MADS domain proteins of other streptophyte algae were obtained from the corresponding
1306 genome annotations and transcriptomic data (OneThousandPlantTranscriptomesInitiative, 2019).
1307 Additional MADS-domain proteins of Zygnematophyceae were identified by BLAST against
1308 transcriptome data available at NCBI's sequence read archive (SRA) (Sayers et al., 2021).
1309 MADS-domain-protein sequences were aligned using MAFFT v7.310 (Katoh and Standley,
1310 2013) with default options. Sequences with bad fit to the MADS domain were excluded and the
1311 remaining sequences realigned, and trimmed using trimAl (Capella-Gutierrez et al., 2009) with
1312 options “-gt .9 -st .0001”. A maximum likelihood phylogeny was reconstructed using
1313 RAxML v8.2.12 (Stamatakis 2014) on the CIPRES Science Gateway (Miller et al., 2011).

1314

1315 **LC-MS/MS analysis of abscisic acid**

1316 Abscisic acid was determined in *Physcomitrium patens* samples using the LC-MS/MS system
1317 which consisted of Nexera X2 UPLC (Shimadzu) coupled QTRAP 6500+ mass spectrometer
1318 (Sciex). Chromatographic separations were carried out using the Acclaim RSLC C18 column
1319 (150×2.1 mm, 2.2µm, Thermo Scientific) employing acetonitrile/water+0.1% acetic acid linear
1320 gradient. The mass spectrometer was operated in negative ESI mode. Data was acquired in MRM
1321 mode using following transitions: 1) ABA 263.2->153.1 (-14 eV), 263.2->219.1 (-18 eV); 2)
1322 ABA -D6 (IS) 269.2->159.1 (-14 eV), 269.2->225.1 (-18 eV); declustering potential was -45 V.
1323 Freeze-dried moss samples were ground using the metal beads in homogenizer (Bioprep-24) to a
1324 fine powder. Accurately weighted (about 20mg) samples were spiked with isotopically labeled
1325 ABA -D6 (total added amount was 2 ng) and extracted with 1.5 ml acetonitrile/water (1:1)
1326 solution acidified with 0.1% formic acid. Extraction was assisted by sonication (Elma S 40 H, 15
1327 min, two cycles) and solution was left overnight for completion of extraction. Liquid was filtered
1328 through 0.2 µm regenerated cellulose membrane filters, evaporated to dryness upon a stream of
1329 dry nitrogen and redissolved in 100 µl extraction solution.

1330

1331 REFERENCES

- 1332 Adamowski M, Friml J. (2015). PIN-dependent auxin transport: action, regulation, and evolution. *Plant*
1333 *Cell*, 27: 20-32, [10.1105/tpc.114.134874](https://doi.org/10.1105/tpc.114.134874)
- 1334 Aigner S, Remias D, Karsten U, Holzinger A 2013. Unusual phenolic compounds contribute to the
1335 ecophysiological performance in the purple-colored green alga *Zygoonidium ericetorum*
1336 (*Zygnematophyceae*, *Streptophyta*) from a high-alpine habitat. *J. Phycol.* 49: 648-660.
- 1337 Antipov, D., Korobeynikov, A., McLean, J.S., and Pevzner, P.A. (2016). HybridSPAdes: An algorithm
1338 for hybrid assembly of short and long reads. *Bioinformatics* 32, 1009–1015.
1339 [10.1093/bioinformatics/btv688](https://doi.org/10.1093/bioinformatics/btv688).
- 1340 Arc E, Pichrtová M, Kranner I, Holzinger A 2020. Pre-akinete formation in *Zygnema* sp. from polar
1341 habitats is associated with metabolite re-arrangement. *J Exp. Bot.* 71: 3314–3322.
1342 [doi:10.1093/jxb/eraa123](https://doi.org/10.1093/jxb/eraa123)
- 1343 Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., DePamphilis, C., Albert, V.A.,
1344 Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The selaginella genome identifies genetic
1345 changes associated with the evolution of vascular plants. *Science* 332, 960–963.
1346 [10.1126/science.1203810](https://doi.org/10.1126/science.1203810)
- 1347 Bassil, E., Ohto, M., Esumi, T., Tajima, H., Zhu, Z., Cagnac, O., Belmonte, M., Peleg, Z., Yamaguchi, T.,
1348 and Blumwald, E. (2011). The *Arabidopsis* Intracellular Na⁺/H⁺ Antiporters NHX5 and NHX6 Are
1349 Endosome Associated and Necessary for Plant Growth and Development. *The Plant Cell* 23, 224–
1350 239. <https://doi.org/10.1105/tpc.110.079426>.

- 1351 Becker B, Feng X, Yin Y, Holzinger A 2020. Desiccation tolerance in streptophyte green algae and the
1352 algae to land plant transition: evolution of LEA and MIP protein families within the viridiplantae. *J*
1353 *Exp. Bot.* 71: 3270–3278. doi: 10.1093/jxb/eraa105
- 1354 Bergman, C.M., and Quesneville, H. (2007). Discovering and detecting transposable elements in genome
1355 sequences. *Briefings in Bioinformatics* 8, 382–392. 10.1093/bib/bbm048.
- 1356 Bienz, M. (2006) The PHD finger, a nuclear protein-interaction domain. *Trends Biochem. Sci.* 31(1):35-
1357 40. doi: 10.1016/j.tibs.2005.11.001
- 1358 Bischoff, H. W., Bold, H.C. (1963) *Phycological studies IV. Some soil algae from enchanted rock and*
1359 *related algal species.* University of Texas Publication No. 6318. p. 1-95. Bräutigam, A., Kajala, K.,
1360 Wullenweber, J., Sommer, M., Gagneul, D., Weber, K.L., Carr, K.M., Gowik, U., Maß, J., Lercher,
1361 M.J., et al. (2011). An mRNA Blueprint for C4 Photosynthesis Derived from Comparative
1362 Transcriptomics of Closely Related C3 and C4 Species. *Plant Physiology* 155, 142–156.
1363 <https://doi.org/10.1104/pp.110.159442>.
- 1364 Bjornson, M., Zipfel, C. (2021) Plant immunity: Crosstalk between plant immune receptors. *Curr. Biol.*
1365 31:PR796-R798. <https://doi.org/10.1016/j.cub.2021.04.080>
- 1366 Blaby, I.K., Blaby-Haas, C.E., Tourasse, N., Hom, E.F.Y., Lopez, D., Aksoy, M., Grossman, A., Umen,
1367 J., Dutcher, S., Porter, M., et al. (2014). The *Chlamydomonas* genome project: a decade on. *Trends*
1368 *in Plant Science* 19, 672–680. 10.1016/j.tplants.2014.05.008
- 1369 Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D.D., Gurnon, J., Ladunga, I.,
1370 Lindquist, E., Lucas, S., et al. (2012). The genome of the polar eukaryotic microalga *Coccomyxa*
1371 *subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* 13, R39. 10.1186/gb-2012-13-5-r39
- 1372 Blanton, R.L., Fuller, D., Iranfar, D., Loomis, W.F. (2000) The cellulose synthase gene of *Dictyostelium*.
1373 *Proc. Natl. Acad. Sci. USA* 97(5):2391-2396. <https://doi.org/10.1073/pnas.040565697>
- 1374 Blum, M., Boehler, M. Rrandall, E., Young, V., Csukai, M., Kraus, S., Moulin, F., Scalliet, G. Avrova,
1375 A.O, Whisson, S.C., Fonne-Pfister, R. (2010) Mandipropamid targets the cellulose synthase-like
1376 PiCesA3 to inhibit cell wall biosynthesis in the oomycete plant pathogen, *Phytophthora infestans*.
1377 *Mol. Plant Pathol.* 11(2):227-243 <https://doi.org/10.1111/j.1364-3703.2009.00604.x>
- 1378 Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence
1379 data. *Bioinformatics* 30, 2114–2120. 10.1093/bioinformatics/btu170.
- 1380 Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R.,
1381 Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the
1382 *Marchantia polymorpha* Genome. *Cell* 171, 287-299.e15. 10.1016/j.cell.2017.09.030
- 1383 Busch A, Hess S (2021) Sunscreen mucilage: a photoprotective adaptation found in terrestrial green algae
1384 (*Zygnematophyceae*) *Eur J Phycol* DOI: 10.1080/09670262.2021.1898677
- 1385 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009)
1386 BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421
- 1387 Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using
1388 MAKER and MAKER-P. *CP in Bioinformatics* 48. 10.1002/0471250953.bi0411s48.
- 1389 Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T.(2009) trimAl: a tool for automated alignment
1390 trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973,
1391 <https://doi.org/10.1093/bioinformatics/btp348>
- 1392 Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated
1393 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973.
1394 10.1093/bioinformatics/btp348.

- 1395 Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in
1396 Phylogenetic Analysis. *Molecular Biology and Evolution* 17, 540–552.
1397 [10.1093/oxfordjournals.molbev.a026334](https://doi.org/10.1093/oxfordjournals.molbev.a026334).
- 1398 Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T. Sun, W., Li, X., Xu, Y., Zhang, Y., Wittek, S.,
1399 Reder, T., Günther, G., Gontcharov, A., Wang, S., Li, L., Liu, X., Wang, J., Yang, H., Xu, X.,
1400 Delaux, P.M., Melkonian, B., Wong, G.K. & Melkonian, M. (2019). Genomes of subaerial
1401 Zygnematophyceae provide insights into land plant evolution. *Cell* 179: 1057–1067
- 1402 Chikhi, R., and Medvedev, P. (2014). Informed and automated k -mer size selection for genome
1403 assembly. *Bioinformatics* 30, 31–37. [10.1093/bioinformatics/btt310](https://doi.org/10.1093/bioinformatics/btt310).
- 1404 Crescente, J.M., Zavallo, D., Helguera, M., and Vanzetti, L.S. (2018). MITE Tracker: an accurate
1405 approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC*
1406 *Bioinformatics* 19, 348. [10.1186/s12859-018-2376-y](https://doi.org/10.1186/s12859-018-2376-y).
- 1407 da Cunha, L., McFall, A.J., and Mackey, D. (2006). Innate immunity in plants: a continuum of layered
1408 defenses. *Microbes and Infection* 8, 1372–1381. [10.1016/j.micinf.2005.12.018](https://doi.org/10.1016/j.micinf.2005.12.018)
- 1409 De Clerck, O., Kao, S.-M., Bogaert, K.A., Blomme, J., Foflonker, F., Kwantes, M., Vancaester, E.,
1410 Vanderstraeten, L., Aydogdu, E., Boesger, J., et al. (2018). Insights into the Evolution of
1411 Multicellularity from the Sea Lettuce Genome. *Current Biology* 28, 2921-2933.e5.
1412 [10.1016/j.cub.2018.08.015](https://doi.org/10.1016/j.cub.2018.08.015).
- 1413 de Vries, J., Curtis, B.A., Gould, S.B., and Archibald, J.M. (2018). Embryophyte stress signaling evolved
1414 in the algal progenitors of land plants. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3471–E3480.
1415 <https://doi.org/10.1073/pnas.1719230115>.
- 1416 de Vries, J., de Vries, S., Curtis, B.A., Zhou, H., Penny, S., Feussner, K., Pinto, D.M., Steinert, M.,
1417 Cohen, A.M., von Schwatzenberg, K., Archibald, J.M. (2020). Heat stress response in the closest
1418 algal relatives of land plants reveals conserved stress signaling circuits. *Plant J.* 103: 1025-1048.
1419 <https://doi.org/10.1111/tpj.14782>
- 1420 Delaux, P.-M., Xie, X., Timme, R.E., Puech-Pages, V., Dunand, C., Lecompte, E., Delwiche, C.F.,
1421 Yoneyama, K., Bécard, G., and Séjalon-Delmas, N. (2012). Origin of strigolactones in the green
1422 lineage. *New Phytol* 195, 857–871. [10.1111/j.1469-8137.2012.04209.x](https://doi.org/10.1111/j.1469-8137.2012.04209.x).
- 1423 Delwiche, C.F. and Cooper, E.D. (2015). The Evolutionary Origin of a Terrestrial Flora. *Curr. Biol.* 25:
1424 R899-910
- 1425 Di Genova, A., Buena-Atienza, E., Ossowski, S., and Sagot, M.-F. (2021). Efficient hybrid de novo
1426 assembly of human genomes with WENGAN. *Nat Biotechnol* 39, 422–430. [10.1038/s41587-020-](https://doi.org/10.1038/s41587-020-00747-w)
1427 [00747-w](https://doi.org/10.1038/s41587-020-00747-w).
- 1428 Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: de novo assembly of organelle
1429 genomes from whole genome data. *Nucleic Acids Res*, gkw955. [10.1093/nar/gkw955](https://doi.org/10.1093/nar/gkw955).
- 1430 Domozych, D.S., Bagdan, K. (2022) The cell biology of Charophytes: Exploring the past and models for
1431 the future. *Plant Physiology* 190: 1588–1608. <https://doi.org/10.1093/plphys/kiac390>.
- 1432 Dong J, Hudson ME. (2022) WII2_{Rhgl} interacts with DELLAs and mediates soybean cyst nematode
1433 resistance through hormone pathways. *Plant Biotechnol J.* 20, 283-296. doi: [10.1111/pbi.13709](https://doi.org/10.1111/pbi.13709).
- 1434 Dragwidge, J.M., Ford, B.A., Ashnest, J.R., Das, P., and Gendall, A.R. (2018). Two Endosomal NHX-
1435 Type Na⁺/H⁺ Antiporters are Involved in Auxin-Mediated Development in *Arabidopsis thaliana*.
1436 *Plant and Cell Physiology* 59, 1660–1669. <https://doi.org/10.1093/pcp/pcy090>.

- 1437 Dure L, 3rd, Greenway SC, Galau GA (1981) Developmental biochemistry of cottonseed embryogenesis
1438 and germination: Changing messenger ribonucleic acid populations as shown by in vitro and in vivo
1439 protein synthesis. *Biochemistry* 20: 4162–4168.
- 1440 Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755-63. doi:
1441 10.1093/bioinformatics/14.9.755.
- 1442 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput.
1443 *Nuc. Acids Res.* 32(5):1792–1797, <https://doi.org/10.1093/nar/gkh340>
- 1444 Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative
1445 genomics. *Genome Biology* 20, 238. 10.1186/s13059-019-1832-y.
- 1446 Faik, A., Bar-Peled, M., DeRocher, A.E., Zeng, W., Perrin, R.M., Wilkerson, C., Raikhel, N.V., and
1447 Keegstra, K. (2000). Biochemical Characterization and Molecular Cloning of an α -1,2-
1448 Fucosyltransferase That Catalyzes the Last Step of Cell Wall Xyloglucan Biosynthesis in Pea.
1449 *Journal of Biological Chemistry* 275, 15082–15089. <https://doi.org/10.1074/jbc.M000677200>.
- 1450 Feng X, Holzinger A, Permann C, Anderson D, Yin Y 2021. Characterization of two *Zygnema* strains
1451 (SAG 698-1a and 698-1b) and a rapid method to estimate genome size of Zygnematophycean green
1452 algae. *Frontiers Plant Science*. doi:10.3389/fpls.2021.610381
- 1453 Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity
1454 searching. *Nucleic Acids Research* 39, W29–W37. 10.1093/nar/gkr367
- 1455 Fitzek E, Orton L, Entwistle S, Grayburn WS, Ausland C, Duvall MR, Yin Y. 2019. Cell wall enzymes in
1456 *Zygnema circumcarinatum* UTEX 1559 respond to osmotic stress in a plant-like fashion. *Frontiers in*
1457 *Plant Science* 10, 732.
- 1458 Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jürgens G. 2003. Efflux-
1459 dependent auxin gradients establish the apical–basal axis of Arabidopsis. *Nature* 426, 147–153.
- 1460 Functional Characterization of a Glycosyltransferase from the Moss *Physcomitrella patens* Involved in the
1461 Biosynthesis of a Novel Cell Wall Arabinoglucan
- 1462 Gauch, H. G. (1966). Studies on the life cycle and genetics of *Zygnema*. [Master’s Thesis]. Ithaca, NY:
1463 Cornell University.
- 1464 Glantz, S.T., Carpenter, E.J., Melkonian, M., Gardner, K.H., Boyden, E.S., Wong, G.K.-S., and Chow,
1465 B.Y. (2016). Functional and topological diversity of LOV domain photoreceptors. *Proc. Natl. Acad.*
1466 *Sci. U.S.A.* 113. <https://doi.org/10.1073/pnas.1509428113>.
- 1467 Gontcharov AA (2008) Phylogeny and classification of Zygnematophyceae (Streptophyta): current state
1468 of affairs. *Fottea* 8(2): 87-104.
- 1469 Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L.,
1470 Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data
1471 without a reference genome. *Nat Biotechnol* 29, 644–652. 10.1038/nbt.1883.
- 1472 Gramzow, L., Ritz, M.S., Theissen, G. (2010) On the origin of MADS-domain transcription factors.
1473 *Trends Genet* 26(4):149-53. doi: 10.1016/j.tig.2010.01.004
- 1474 Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1:
1475 expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47,
1476 W59–W64. 10.1093/nar/gkz238
- 1477 Grigoriev, I.V., Hayes, R.D., Calhoun, S., Kamel, B., Wang, A., Ahrendt, S., Dusheyko, S., Nikitin, R.,
1478 Mondo, S.J., Salamov, A., et al. (2021). PhycoCosm, a comparative algal genomics resource.
1479 *Nucleic Acids Research* 49, D1004–D1011. 10.1093/nar/gkaa898.

- 1480 Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New
1481 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance
1482 of PhyML 3.0. *Syst. Biol.* 59(3):307–321.
- 1483 Haas, B.J. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment
1484 assemblies. *Nucleic Acids Research* 31, 5654–5666. [10.1093/nar/gkg770](https://doi.org/10.1093/nar/gkg770).
- 1485 Han JW, Kim GH (2013) An ELIP-like gene in the freshwater green alga, *Spirogyra varians*
1486 (*Zygnematales*), is regulated by cold stress and CO₂ influx. *J Appl Phycol* 25: 1297–1307.
- 1487 Herburger K, Remias D, Holzinger A 2016. The green alga *Zygonium ericetorum* (*Zygnematophyceae*,
1488 Charophyta) shows high iron and aluminium tolerance: Protection mechanisms and photosynthetic
1489 performance. *FEMS Microbiol Ecol*, 92: fiw103, doi: [10.1093/femsec/fiw103](https://doi.org/10.1093/femsec/fiw103)
- 1490 Herburger K, Ryan L, Popper ZA, Holzinger A 2018. Localisation and substrate specificities of
1491 transglycanases in charophyte algae relate to development and morphology *J. Cell Sci.* 131:
1492 [jcs203208](https://doi.org/10.1242/jcs.203208) doi: [10.1242/jcs.203208](https://doi.org/10.1242/jcs.203208). Holzinger A, Albert A, Aigner S, Uhl J, Schmitt-Kopplin P,
1493 Trumhová K, Pichrtová M 2018. Arctic, Antarctic and temperate green algae *Zygnema* spp. under
1494 UV-B stress: Vegetative cells perform better than pre-akinetes. *Protoplasma* 255: 1239-1252.
1495 doi:[10.1007/s00709-018-1225-1](https://doi.org/10.1007/s00709-018-1225-1)
- 1496 Herburger K, Xin A, Holzinger A 2019. Homogalacturonan accumulation in cell walls of the green alga
1497 *Zygnema* sp. (Charophyta) increases desiccation resistance. *Frontiers Plant Science* doi:
1498 [10.3389/fpls.2019.00540](https://doi.org/10.3389/fpls.2019.00540)
- 1499 Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S. (2018) UFBBoot2: Improving the
1500 Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35(2): 518–522. Doi:[10.1093/molbev/msx281](https://doi.org/10.1093/molbev/msx281)
- 1501 Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori,
1502 H., Tajima, N., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant
1503 terrestrial adaptation. *Nat Commun* 5, 3978. [10.1038/ncomms4978](https://doi.org/10.1038/ncomms4978).
- 1504 Hutin, et al. (2003) Early light-induced proteins protect Arabidopsis from photooxidative stress. *Proc Natl*
1505 *Acad Sci USA* 100, 4921–4926
- 1506 Hwang, I., and Sheen, J. (2001). Two-component circuitry in Arabidopsis cytokinin signal
1507 transduction. *Nature* 413, 383–389. <https://doi.org/10.1038/35096500>.
- 1508 Jacob, P., Kim, N.H., Wu, F., El-Kasbi, F., Chi, Y., Walton, W.G., Furzer, O.J., Lietzan, A.D., Sunil, S.,
1509 Kempthorn, K., Redinbo, M.R., Pei, Z.-M., Wan, L., Dangl, J.L. (2021) Plant "helper" immune
1510 receptors are Ca²⁺-permeable nonselective cation channels. *Science* 373(6553):420-425. doi:
1511 [10.1126/science.abg7917](https://doi.org/10.1126/science.abg7917).
- 1512 Jia, K.-P., Mi, J., Ali, S., Ohyanagi, H., Moreno, J.C., Ablazov, A., Balakrishna, A., Berqdar, L., Fiore,
1513 A., Diretto, G., et al. (2022). An alternative, zeaxanthin epoxidase-independent abscisic acid
1514 biosynthetic pathway in plants. *Molecular Plant* 15, 151–166. [10.1016/j.molp.2021.09.008](https://doi.org/10.1016/j.molp.2021.09.008).
- 1515 Jiao, C., Sørensen, I., Sun, X., Sun, H., Behar, H., Alseikh, S., Philippe, G., Palacio Lopez, K., Sun, L.,
1516 Reed, R., et al. (2020). The *Penium margaritaceum* Genome: Hallmarks of the Origins of Land
1517 Plants. *Cell* 181, 1097-1111.e12. <https://doi.org/10.1016/j.cell.2020.04.019>.
- 1518 Johnson, L.S., Eddy, S.R., Portugaly E. (2010) Hidden Markov model speed heuristic and iterative HMM
1519 search procedure. *BMC Bioinformatics* 11: 431
- 1520 Kajitani, R., Yoshimura, D., Okuno, M., Minakuchi, Y., Kagoshima, H., Fujiyama, A., Kubokawa, K.,
1521 Kohara, Y., Toyoda, A., and Itoh, T. (2019). *Platanus-allee* is a de novo haplotype assembler
1522 enabling a comprehensive access to divergent heterozygous regions. *Nat Commun* 10, 1702.
1523 [10.1038/s41467-019-09575-2](https://doi.org/10.1038/s41467-019-09575-2).

- 1524 Kalvari, I., Nawrocki, E.P., Argasinska, J., Quinones-Olvera, N., Finn, R.D., Bateman, A., and Petrov,
1525 A.I. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in*
1526 *Bioinformatics* 62, e51. [10.1002/cpbi.51](https://doi.org/10.1002/cpbi.51).
- 1527 Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jeremiin, L.S. (2017) ModelFinder:
1528 fast model selection for accurate phylogenetic estimates. *Nat. Meth.* 14:587–589.
- 1529 Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7:
1530 Improvements in Performance and Usability. *Molecular Biology and Evolution* 30, 772–780.
1531 [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- 1532 Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz,
1533 D.C., Tanaka, T., Wu, J., Zhou, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare
1534 reference genome using next generation sequence and optical map data. *Rice* 6, 4. [10.1186/1939-](https://doi.org/10.1186/1939-8433-6-4)
1535 [8433-6-4](https://doi.org/10.1186/1939-8433-6-4).
- 1536 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A.,
1537 Markowitz, S., Duran, C., et al. (2012). Geneious Basic: An integrated and extendable desktop
1538 software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
1539 [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).
- 1540 Kerk, D., Templeton, G., and Moorhead, G.B.G. (2008). Evolutionary Radiation Pattern of Novel Protein
1541 Phosphatases Revealed by Analysis of Protein Data from the Completely Sequenced Genomes of
1542 Humans, Green Algae, and Higher Plants. *Plant Physiology* 146, 323–324.
1543 <https://doi.org/10.1104/pp.107.111393>.
- 1544 Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment
1545 and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907–915. [10.1038/s41587-](https://doi.org/10.1038/s41587-019-0201-4)
1546 [019-0201-4](https://doi.org/10.1038/s41587-019-0201-4).
- 1547 Kodama, K., Rich, M.K., Yoda, A., Shimazaki, S., Xie, X., Akiyama, K., Mizuno, Y., Komatsu, A., Luo,
1548 Y., Suzuki, H., et al. (2022). An ancestral function of strigolactones as symbiotic rhizosphere signals.
1549 *Nat Commun* 13, 3974. [10.1038/s41467-022-31708-3](https://doi.org/10.1038/s41467-022-31708-3).
- 1550 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu:
1551 scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation.
1552 *Genome Res.* 27, 722–736. [10.1101/gr.215087.116](https://doi.org/10.1101/gr.215087.116)
- 1553 Krupková, E., Immerzeel, P., Pauly, M., and Schmölling, T. (2007). The TUMOROUS SHOOT
1554 DEVELOPMENT2 gene of *Arabidopsis* encoding a putative methyltransferase is required for cell
1555 adhesion and co-ordinated plant development: The *Arabidopsis* TSD2 gene. *The Plant Journal* 50,
1556 735–750. [10.1111/j.1365-3113X.2007.03123.x](https://doi.org/10.1111/j.1365-3113X.2007.03123.x).
- 1557 Kumari, P., Gupta, A., Yadav, S. (2021). Thioredoxins as Molecular Players in Plants, Pests, and
1558 Pathogens. In: Singh, I.K., Singh, A. (eds) *Plant-Pest Interactions: From Molecular Mechanisms to*
1559 *Chemical Ecology*. Springer, Singapore. https://doi.org/10.1007/978-981-15-2467-7_6
- 1560 Lam, K.C., Ibrahim, R.K., Behdad, B., and Dayanandan, S. (2007). Structure, function, and evolution of
1561 plant O-methyltransferases. *50*, 13.
- 1562 Lang, D., Ullrich, K.K., Murat, F., Fuchs, J., Jenkins, J., Haas, F.B., Piednoel, M., Gundlach, H., Van Bel,
1563 M., Meyberg, R., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss
1564 genome structure and evolution. *The Plant Journal* 93, 515–533. [10.1111/tpj.13801](https://doi.org/10.1111/tpj.13801).
- 1565 Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9,
1566 357–359. [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923).

- 1567 Laslett, D., Canbäck B. (2008) ARWEN, a program to detect tRNA genes in metazoan mitochondrial
1568 nucleotide sequences. *Bioinformatics* 24;172-175
- 1569 Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
1570 display and annotation. *Nucleic Acids Research* 49, W293–W296. 10.1093/nar/gkab301.
- 1571 Letunic, I., Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
1572 display and annotation. *Nuc. Acids Res* 49(W1):W293–W296. Doi:10.1093/nar/gkab301
- 1573 Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N.,
1574 Kuo, L.-Y., Li, Z., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial
1575 symbioses. *Nature Plants* 4, 460–472. 10.1038/s41477-018-0188-8.
- 1576 Li, F.-W., Melkonian, M., Rothfels, C.J., Villarreal, J.C., Stevenson, D.W., Graham, S.W., Wong, G.K.-
1577 S., Pryer, K.M., and Mathews, S. (2015). Phytochrome diversity in green plants and the origin of
1578 canonical plant phytochromes. *Nat Commun* 6, 7852. 10.1038/ncomms8852.
- 1579 Li, F.-W., Villarreal, J.C., Kelly, S., Rothfels, C.J., Melkonian, M., Frangedakis, E., Ruhsam, M., Sigel,
1580 E.M., Der, J.P., Pittermann, J., et al. (2014). Horizontal transfer of an adaptive chimeric
1581 photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6672–6677.
1582 10.1073/pnas.1319929111.
- 1583 Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100.
1584 10.1093/bioinformatics/bty191.
- 1585 Lopez-Obando, M., Conn, C.E., Hoffmann, B., Bythell-Douglas, R., Nelson, D.C., Rameau, C., and
1586 Bonhomme, S. (2016). Structural modelling and transcriptional responses highlight a clade of
1587 PpKAI2-LIKE genes as candidate receptors for strigolactones in *Physcomitrella patens*. *Planta* 243,
1588 1441–1453. 10.1007/s00425-016-2481-y.
- 1589 Lowe, T.M., and Chan, P.P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of
1590 transfer RNA genes. *Nucleic Acids Res* 44, W54–W57. 10.1093/nar/gkw413.
- 1591 Ma, J., Wang, S., Zhu, X., Sun, G., Chang, G., Li, L., Hu, X., Zhang, S., Zhou, Y., Song, C.-P., et al.
1592 (2022). Major episodes of horizontal gene transfer drove the evolution of land plants. *Molecular*
1593 *Plant* 15, 857–871. <https://doi.org/10.1016/j.molp.2022.02.001>.
- 1594 Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018).
1595 MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* 14, e1005944.
1596 10.1371/journal.pcbi.1005944.
- 1597 Marchant, D.B., et al. (2022) Dynamic genome evolution in a model fern. *Nat Plants* 8(9): 1038–1051.
1598 doi: 10.1038/s41477-022-01226-7
- 1599 Matthews, P.R., Schindler, M., Howles, P., Arioli, T., and Williamson, R.E. (2010). A CESA from
1600 *Griffithsia monilis* (Rhodophyta, Florideophyceae) has a family 48 carbohydrate-binding module.
1601 *Journal of Experimental Botany* 61, 4461–4468. 10.1093/jxb/erq254.
- 1602 Matthyse, A.G., Deschet, K. Williams, M., Smith, W.C. (2004) A functional cellulose synthase from
1603 ascidian epidermis. *Proc. Natl. Acad. Sci USA* 101(4):986-991.
1604 <https://doi.org/10.1073/pnas.0303623101>
- 1605 McFadden, G.I., Melkonian, M. (1986): Use of Hepes buffer for microalgal culture media and fixation for
1606 electron microscopy. *Phycologia* 25, 551-557
- 1607 McKay, D.W., McFarlane, H.E., Qu, Y., Situmorang, A., Gilliam, M., and Wege, S. (2022). Plant Trans-
1608 Golgi Network/Early Endosome pH regulation requires Cation Chloride Cotransporter (CCC1).
1609 *ELife* 11, e70701. <https://doi.org/10.7554/eLife.70701>.

- 1610 Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2021). CAFE 5 models variation in
1611 evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518.
1612 10.1093/bioinformatics/btaa1022.
- 1613 Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A.,
1614 Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., et al. (2007). The *Chlamydomonas* genome
1615 reveals the evolution of key animal and plant functions. *Science* 318, 245–251.
1616 10.1126/science.1143609.
- 1617 Mikkelsen, M.D., Harholt, J., Westereng, B., Domozych, D., Fry, S.C., Johansen, I.E., Fangel, J.U.,
1618 Łężyk, M., Feng, T., Nancke, L., et al. (2021). Ancient origin of fucosylated xyloglucan in
1619 charophycean green algae. *Commun Biol* 4, 754. <https://doi.org/10.1038/s42003-021-02277-w>.
- 1620 Miller, M.A., Pfeiffer, W., Schwartz, T. (2011) The CIPRES science gateway: a community resource for
1621 phylogenetic analyses. *TG '11: Proceedings of the 2011 TeraGrid Conference: Extreme Digital
1622 Discovery*, July 2011 Article No.: 41:1–8. Doi:10.1145/2016741.2016785
- 1623 Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear,
1624 R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
1625 Genomic Era. *Mol. Biol. Evol.* 37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- 1626 Missbach, S., Weis, B.L., Martin, R., Simm, S., Bohnsack, M.T., and Schleiff, E. (2013). 40S Ribosome
1627 Biogenesis Co-Factors Are Essential for Gametophyte and Embryo Development. *PLOS ONE* 8, 19.
- 1628 Mitchell, M.C., Meyer, M.T., and Grif, H. (2014). Dynamics of Carbon-Concentrating Mechanism
1629 Induction and Protein Relocalization during the Dark-to-Light Transition in Synchronized
1630 *Chlamydomonas reinhardtii*[W][OPEN]. *166*, 10.
- 1631 Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Bel, M.V., Poulain, J.,
1632 Katinka, M., Hohmann-Marriott, M.F., et al. (2012). Gene functionalities and genome structure in
1633 *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome
1634 Biology* 13, R74.
- 1635 Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective
1636 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and
1637 Evolution* 32, 268–274. 10.1093/molbev/msu300.
- 1638 Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B.,
1639 Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The *Chara* Genome: Secondary Complexity
1640 and Implications for Plant Terrestrialization. *Cell* 174, 448–464.e24. 10.1016/j.cell.2018.06.033.
- 1641 Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., Vezzi, F., Delhomme,
1642 N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer
1643 genome evolution. *Nature* 497, 579–584. 10.1038/nature12211.
- 1644 One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics
1645 of green plants. *Nature* 574, 679–685 (2019). <https://doi.org/10.1038/s41586-019-1693-2>
- 1646 Palacio-López K, Tinaz B, Holzinger A, Domozych DS 2019. Arabinogalactan proteins and the
1647 extracellular matrix of charophytes: a sticky business. *Frontiers Plant Science* doi:
1648 10.3389/fpls.2019.00447
- 1649 Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R.,
1650 Derelle, E., Rombauts, S., et al. (2007). The tiny eukaryote *Ostreococcus* provides genomic insights
1651 into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 7705–7710

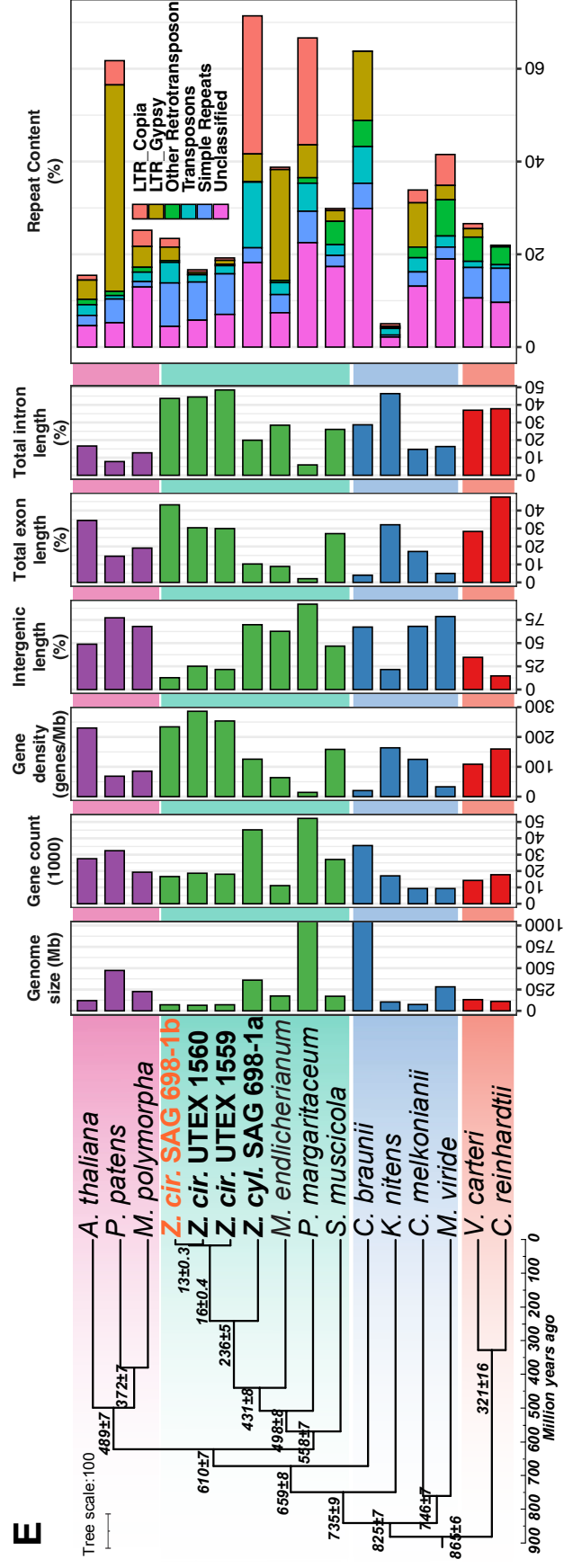
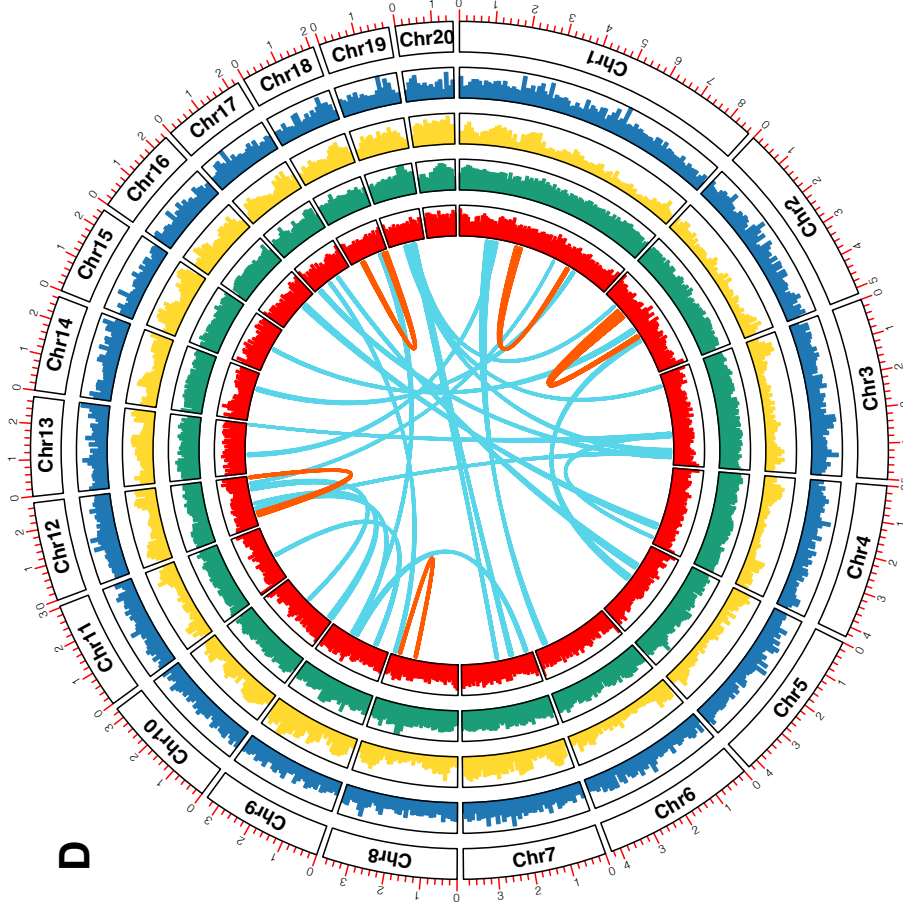
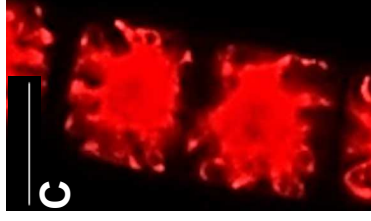
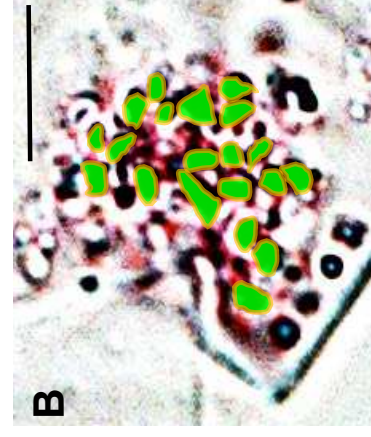
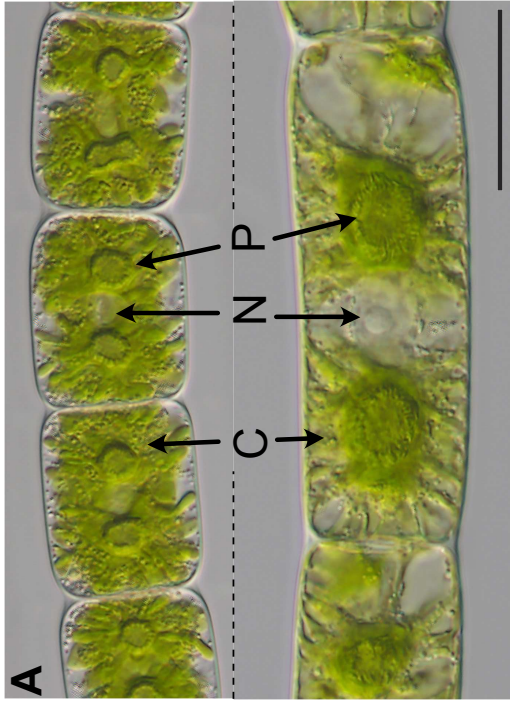
- 1652 Permann C, Becker B, Holzinger A 2022b. Temperature- and light stress adaptations in
1653 Zygnematophyceae: The challenges of a semi-terrestrial lifestyle. *Frontiers Plant Science*.
1654 13:945394. doi: 10.3389/fpls.2022.945394.
- 1655 Permann C, Gierlinger N, Holzinger A 2022a. Zygosporangia of the green alga *Spirogyra*: new insights from
1656 structural and chemical imaging. *Frontiers Plant Science*. 13:1080111. doi:
1657 10.3389/fpls.2022.1080111
- 1658 Permann C, Herburger K, Niedermeier M, Felhofer M, Gierlinger N, Holzinger A 2021. Cell wall
1659 characteristics during sexual reproduction of *Mougeotia* sp. (Zygnematophyceae) revealed by
1660 electron microscopy, comprehensive polymer profiling and RAMAN spectroscopy. *Protoplasma*
1661 258: 1261–1275. doi: 10.1007/s00709-021-01659-5
- 1662 Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression
1663 analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11, 1650–1667.
1664 10.1038/nprot.2016-095.
- 1665 Pfeifer L, Utermöhlen J, Happ K, Permann C, Holzinger A, von Schwartzberg K, Classen B 2022.
1666 Search for evolutionary roots of land plant arabinogalactan-proteins in charophytes: presence of a
1667 rhamnogalactan-protein in *Spirogyra pratensis* (Zygnematophyceae). *Plant J*. 109: 568–584. doi:
1668 10.1111/tbj.15577
- 1669 Pichrtová M, Holzinger A, Kulichová J, Ryšánek D, Šoljaková T, Trumhová K, Nemcová Y 2018.
1670 Molecular and morphological diversity of *Zygnema* and *Zygnemopsis* (Zygnematophyceae,
1671 Streptophyta) on Svalbard (High Arctic). *Eur. J. Phycol* 53: 492-508. doi:
1672 10.1080/09670262.2018.1476920
- 1673 Pichrtová M, Remias D, Lewis LA, Holzinger A 2013. Changes in phenolic compounds and cellular
1674 ultrastructure of arctic and Antarctic strains of *Zygnema* (Zygnematales, Streptophyta) after exposure
1675 to experimentally enhanced UV to PAR ratio. *Microb. Ecol.* 65: 68-83
- 1676 Plett, J.M., and Martin, F.M. (2018). Know your enemy, embrace your friend: using omics to understand
1677 how plants respond differently to pathogenic and mutualistic microorganisms. *Plant J* 93, 729–746.
1678 10.1111/tbj.13802.
- 1679 Prasad BN, Godward MB. (1966) Cytological studies in the genus *Zygnema*. *Cytologia* (Tokyo). 31, 375-
1680 91. doi: 10.1508/cytologia.31.375.
- 1681 Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood
1682 Trees for Large Alignments. *PLoS ONE* 5, e9490. 10.1371/journal.pone.0009490.
- 1683 Price, M.N., Dehal, P.S., Arkin, A.P. (2009) FastTree: Computing Large Minimum Evolution Trees with
1684 Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* 26(7):1641–1650
1685 <https://doi.org/10.1093/molbev/msp077>
- 1686 Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A.,
1687 Mitros, T., Fritz-Laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the
1688 multicellular green alga *Volvox carteri*. *Science* 329, 223–226. 10.1126/science.1188800
- 1689 Proost, S., and Mutwil, M. (2018). CoNekT: an open-source framework for comparative genomic and
1690 transcriptomic network analyses. *Nucleic Acids Research* 46, W133–W140. 10.1093/nar/gky336.
- 1691 Proust, H., Hoffmann, B., Xie, X., Yoneyama, K., Schaefer, D.G., Yoneyama, K., Nogué, F., and
1692 Rameau, C. (2011). Strigolactones regulate protonema branching and act as a quorum sensing-like
1693 signal in the moss *Physcomitrella patens*. *Development* 138, 1531–1539. 10.1242/dev.058495.
- 1694 Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E.,
1695 Hart, D., Theißen, G., Silva, S.S., Parcy, F., Dumas, R., Nanao, M., und Zubieta, C. (2014).

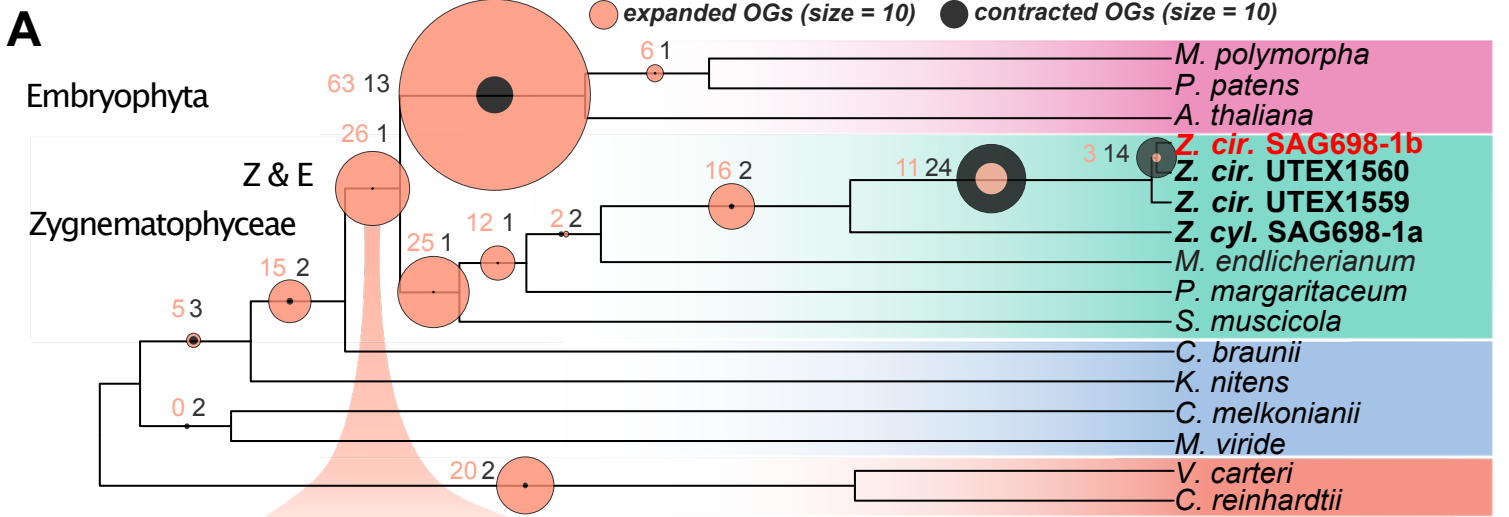
- 1696 Structural basis for the oligomerization of the MADS domain transcription factor SEPALLATA3 in
1697 Arabidopsis. *Plant Cell* 26, 3603-3615.
- 1698 Reddy, A.S.N., Ali, G.S., Celesnik, H. and Day, I.S. (2012) Coping with stresses: roles of calcium- and
1699 calcium/calmodulin-regulated gene expression. *Plant Cell*, 23, 2010–2032.
- 1700 Resemann, H.C., Herrfurth, C., Feussner, K., Hornung, E., Ostendorf, A.K., Gömann, J., Mittag, J., van
1701 Gessel, N., de Vries, J., Ludwig-Müller, J., et al. (2021). Convergence of sphingolipid desaturation
1702 across over 500 million years of plant evolution. *Nature Plants* 7, 219–232.
- 1703 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things):
1704 phytools: R package. *Methods in Ecology and Evolution* 3, 217–223. [10.1111/j.2041-](https://doi.org/10.1111/j.2041-210X.2011.00169.x)
1705 [210X.2011.00169.x](https://doi.org/10.1111/j.2041-210X.2011.00169.x).
- 1706 Rieseberg, T.P., Dadras, A., Fürst-Jansen, J.M.R., Dhabalia Ashok, A., Darienko, T., de Vries, S., Irisarri,
1707 I., and de Vries, J. (2023). Crossroads in the evolution of plant specialized metabolism. *Seminars in*
1708 *Cell & Developmental Biology* 134, 37-58. <https://doi.org/10.1016/j.semcdb.2022.03.004>
- 1709 Rippin M, Becker B, Holzinger A 2017. Enhanced desiccation tolerance in mature cultures of the
1710 streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant Cell Physiol.*
1711 58: 2067-2084. doi: 10.1093/pcp/pcx136
- 1712 Rippin M, Pichrtová M, Arc E, Kranner I, Becker B, Holzinger A 2019. Metatranscriptomic and
1713 metabolite profiling reveals vertical heterogeneity within a *Zygnema* green algal mat from Svalbard
1714 (High Arctic). *Environ Microbiol* 21: 4283-4299. doi:10.1111/1462-2920.14788
- 1715 Roberts, A.W., Lahnstein, J., Hsieh, Y.S.Y., Xing, X., Yap, K., Chaves, A.M., Scavuzzo-Duggan, T.R.,
1716 Dimitroff, G., Lonsdale, A., Roberts E., Bulone, V., Fincher, G.B., Doblin, M.S., Bacic, A., Burton,
1717 R.A. (2018) Functional Characterization of a Glycosyltransferase from the Moss *Physcomitrella*
1718 *patens* Involved in the Biosynthesis of a Novel Cell Wall Arabinoglucan. *Plant Cell* 30:1293–1308,
1719 <https://doi.org/10.1105/tpc.18.00082>
- 1720 Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat Methods* 17, 155–
1721 158. [10.1038/s41592-019-0669-3](https://doi.org/10.1038/s41592-019-0669-3).
- 1722 Rümpler, F., Tessari, Ch., Gramzow, L., Gafert, Ch., Blohs, M. and Theißen, G. (2022). The origin of
1723 floral quartet formation - Ancient exon duplications shaped the evolution of MIKC-type MADS-
1724 domain transcription factor interactions. preprint at bioRxiv
1725 <https://biorxiv.org/cgi/content/short/2022.12.23.521771v1>
- 1726 Samar, D., Kieler, J.B., Klutts, J.S. (2015) Identification and Deletion of Tft1, a Predicted
1727 Glycosyltransferase Necessary for Cell Wall β -1,3;1,4-Glucan Synthesis in *Aspergillus fumigatus*.
1728 *PLoS ONE* 10(2): e0117336. <https://doi.org/10.1371/journal.pone.0117336>
- 1729 Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke
1730 W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala
1731 SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD, Sherry ST. Database
1732 resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2021 Jan
1733 8;49(D1):D10-D17. doi: 10.1093/nar/gkaa892. PMID: 33095870; PMCID: PMC7778943.
- 1734 Scheres, B., and van der Putten, W.H. (2017). The plant percepton connects environment to
1735 development. *Nature* 543, 337–345. <https://doi.org/10.1038/nature22010>.
- 1736 Sekimoto, H., Komiya, A., Tsuyuki, N., Kawai, J., Kanda, N., Ootsuki, R., Suzuki, Y., Toyoda, A.,
1737 Fujiyama, A., Kasahara, M., et al. (2023). A divergent RWP□RK transcription factor determines
1738 mating type in heterothallic *Closterium*. *New Phytologist*, nph.18662. [10.1111/nph.18662](https://doi.org/10.1111/nph.18662).

- 1739 Seppey, M., Manni, M., Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation
1740 Completeness. In: Kollmar, M. (eds) Gene Prediction. Methods in Molecular Biology, vol 1962.
1741 Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9173-0_14
1742 Shinohara, N. Nishitani, K. (2021) Cryogenian Origin and Subsequent Diversification of the Plant Cell-
1743 Wall Enzyme XTH Family. *Plant Cell Physiol.* 62:1874–1889. <https://doi.org/10.1093/pcp/pcab093>
1744 Skokan R, Medvecká E, Viaene T, et al. 2019. PIN-driven auxin transport emerged early in streptophyte
1745 evolution. *Nature Plants* 5, 1114–1119.
1746 Smýkalová, I., Ludvíková, M., Ondráčková, E., Klejdus, B., Bonhomme, S., Kronusová, O., Soukup, A.,
1747 Rozmoš, M., Guzow-Krzemińska, B., and Matúšová, R. (2017). Green microalga *Trebouxia* sp.
1748 produces strigolactone-related compounds. *Biorxiv preprint* 10.1101/195883.
1749 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1750 phylogenies. *Bioinformatics* 30, 1312–1313. 10.1093/bioinformatics/btu033.
1751 Stamenkovic M, Hanelt D (2016) Geographic Geographic distribution and ecophysiological adaptations
1752 of desmids (Zygnematophyceae, Streptophyta) in relation to PAR, UV radiation and temperature: a
1753 review *Hydrobiologia* DOI 10.1007/s10750-016-2958-5
1754 Stancheva et al. (2012) Systematics of the genus *Zygnema* (Zygnematophyceae, Charophyta) from
1755 Californian watersheds. *Journal of Phycology* 48:409–422
1756 Stastny (2008) Desmids from ephemeral pools and aerophytic habitats from the Czech Republic. *Biologia*
1757 63, 888–894
1758 Stiti, N., Giarola, V., and Bartels, D. (2021). From algae to vascular plants: The multistep evolutionary
1759 trajectory of the ALDH superfamily towards functional promiscuity and the emergence of structural
1760 characteristics. *Environmental and Experimental Botany* 185, 104376.
1761 <https://doi.org/10.1016/j.envexpbot.2021.104376>.
1762 Suetsugu, N., Mittmann, F., Wagner, G., Hughes, J., and Wada, M. (2005). A chimeric photoreceptor
1763 gene, NEOCHROME, has arisen twice during plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102,
1764 13705–13709. 10.1073/pnas.0504734102.
1765 Sun, Y., Ben Harpazi, Wijerathna-Yapa, A., Merilo, E., de Vries, J., Michaeli, D., Gal, M., Cuming, A.C.,
1766 Kollist, H., and Mosquna, A. (2019). A ligand-independent origin of abscisic acid perception. *PNAS*
1767 116, 24892–24899. <https://doi.org/10.1073/pnas.1914480116>.
1768 Tanabe, Y., Hasebe, M., Sekimoto, H., Nishiyama, T., Kitani, M., Henschel, K., Münster, T., Theissen,
1769 G., Nozaki, H. and Ito, M. (2005). Characterization of MADS-box genes in charophycean green
1770 algae and its implication for the evolution of MADS-box genes. *Proc. Natl. Acad. Sci. USA* 102,
1771 2436-2441.
1772 Telagathoti, A., Probst, M., Peintner, U. (2021) Habitat, Snow-Cover and Soil pH, Affect the Distribution
1773 and Diversity of Mortierellaceae Species and Their Associations to Bacteria. *Front. Microbiol.* 12
1774 <https://doi.org/10.3389/fmicb.2021.669784>
1775 Theißen, G., Melzer, R. and Rümpler, F. (2016). MADS-domain transcription factors and the floral
1776 quartet model of flower development: linking plant development and evolution. *Development* 143,
1777 3259-3271.
1778 Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017).
1779 GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* 45, W6–
1780 W11. 10.1093/nar/gkx391.

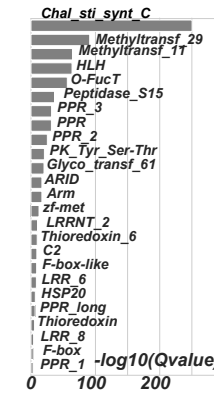
- 1781 Transeau, E.N. (1951). The Zygnemataceae (fresh-water conjugate algae) with keys for the identification
1782 of genera and species, and seven hundred eighty-nine illustrations. pp. 1-327, 789 figs. Columbus:
1783 The Ohio State University Press.
- 1784 Tryfona, T., Theys, T.E., Wagner, T., Stott, K., Keegstra, K., and Dupree, P. (2014). Characterisation of
1785 FUT4 and FUT6 α -(1 \rightarrow 2)-Fucosyltransferases Reveals that Absence of Root Arabinogalactan
1786 Fucosylation Increases Arabidopsis Root Growth Salt Sensitivity. PLoS ONE 9, e93291.
1787 <https://doi.org/10.1371/journal.pone.0093291>.
- 1788 Tsekos, I. (2002) The sites of cellulose synthesis in algae: Diversity and evolution of cellulose-
1789 synthesizing enzyme complexes. J. Phycol. 35:635-655 [mhttps://doi.org/10.1046/j.1529-](https://doi.org/10.1046/j.1529-8817.1999.3540635.x)
1790 [8817.1999.3540635.x](https://doi.org/10.1046/j.1529-8817.1999.3540635.x)
- 1791 Van de Poel, B., Cooper, E.D., Van Der Straeten, D., Chang, C., and Delwiche, C.F. (2016).
1792 Transcriptome Profiling of the Green Alga *Spirogyra pratensis*(Charophyta) Suggests an Ancestral
1793 Role for Ethylene in Cell Wall Metabolism, Photosynthesis, and Abiotic Stress Responses. PLANT
1794 PHYSIOLOGY 172, 533–545. <https://doi.org/10.1104/pp.16.00299>.
- 1795 Vandebussche, F., Yu, N., Li, W., Vanhaelewyn, L., Hamshou, M., Van Der Straeten, D., and Smaghe,
1796 G. (2018). An ultraviolet B condition that affects growth and defense in Arabidopsis. Plant Science
1797 268, 54–63. [10.1016/j.plantsci.2017.12.005](https://doi.org/10.1016/j.plantsci.2017.12.005).
- 1798 Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly
1799 from long uncorrected reads. Genome Res. 27, 737–746. [10.1101/gr.214270.116](https://doi.org/10.1101/gr.214270.116).
- 1800 Vosolsobě S, Skokan R, Petrášek J (2020). The evolutionary origins of auxin transport: what we know
1801 and what we need to know. J Exp. Bot. 71: 3287-3295 doi: [10.1093/jxb/eraa169](https://doi.org/10.1093/jxb/eraa169)
- 1802 Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M.
1803 C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics, 33,
1804 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- 1805 Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q.,
1806 Wortman, J., Young, S.K., et al. (2014). Pilon: An Integrated Tool for Comprehensive Microbial
1807 Variant Detection and Genome Assembly Improvement. PLoS ONE 9, e112963.
1808 [10.1371/journal.pone.0112963](https://doi.org/10.1371/journal.pone.0112963).
- 1809 Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit
1810 Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics &
1811 Bioinformatics 8, 77–80. [10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3).
- 1812 Wang, H.-C., Minh, B.Q., Susko, E., and Roger, A.J. (2018). Modeling Site Heterogeneity with Posterior
1813 Mean Site Frequency Profiles Accelerates Accurate Phylogenomic Estimation. Systematic Biology
1814 67, 216–235. [10.1093/sysbio/syx068](https://doi.org/10.1093/sysbio/syx068).
- 1815 Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y., Xian, W., Song, B., Liang, H., Cheng, S., et al.
1816 (2020). Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. Nat.
1817 Plants 6, 95–106. [10.1038/s41477-019-0560-3](https://doi.org/10.1038/s41477-019-0560-3).
- 1818 Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T. -h., Jin, H., Marler, B., Guo, H., et
1819 al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and
1820 collinearity. Nucleic Acids Research 40, e49–e49. [10.1093/nar/gkr1293](https://doi.org/10.1093/nar/gkr1293).
- 1821 Wang, Y., Tyler, B.M., and Wang, Y. (2019). Defense and Counterdefense During Plant-Pathogenic
1822 Oomycete Infection. Annu. Rev. Microbiol. 73, 667–696. [10.1146/annurev-micro-020518-120022](https://doi.org/10.1146/annurev-micro-020518-120022)
- 1823 Waters, M.T., Gutjahr, C., Bennett, T., and Nelson, D.C. (2017). Strigolactone Signaling and Evolution.
1824 Annu. Rev. Plant Biol. 68, 291–322. [10.1146/annurev-arplant-042916-040925](https://doi.org/10.1146/annurev-arplant-042916-040925).

- 1825 Weijers D, Wagner D. 2016. Transcriptional responses to the auxin hormone. *Annual Review of Plant*
1826 *Biology* 67, 539–574.
- 1827 Wright, E.S. (2015) DECIPHER: harnessing local sequence context to improve protein multiple sequence
1828 alignment. *BMC Bioinformatics* 16:322
- 1829 Xiao, P., Feng, J.-W., Zhu, X.-T., and Gao, J. (2021). Evolution Analyses of CAMTA Transcription
1830 Factor in Plants and Its Enhancing Effect on Cold-tolerance. *Front. Plant Sci.* 12, 758187.
1831 <https://doi.org/10.3389/fpls.2021.758187>.
- 1832 Xu Y, Wang S, Li L, Sahu SK, Petersen M, Liu X, Melkonian M, Zhang G, Liu H. (2019) Molecular
1833 evidence for origin, diversification and ancient gene duplication of plant subtilases (SBTs). *Sci Rep.*
1834 9, 12485. doi: 10.1038/s41598-019-48664-6.
- 1835 Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
1836 *Evolution* 24, 1586–1591. 10.1093/molbev/msm088.
- 1837 Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino
1838 acid sequences. *Genetics* 141, 16411650.
- 1839 Yee, D., and Goring, D.R. (2009). The diversity of plant U-box E3 ubiquitin ligases: from upstream
1840 activators to downstream target substrates. *Journal of Experimental Botany* 60, 1109–1121.
1841 <https://doi.org/10.1093/jxb/ern369>.
- 1842 Yen SK, Chung MC, Chen PC, Yen HE. (2001) Environmental and developmental regulation of the
1843 wound-induced cell wall protein W112 in the halophyte ice plant. *Plant Physiol.* 127, 517-28.
- 1844 Yin, Y., Johns, M.A., Cao, H., Rupani, M. (2014) A survey of plant and algal genomes and
1845 transcriptomes reveals new insights into the evolution and function of the cellulose synthase
1846 superfamily. *BMC Genomics* 15:260
- 1847 Yin, Y., Huang, J., Xu, Y. (2009) The cellulose synthase superfamily in fully sequenced plants and algae.
1848 *BMC Plant Biology* 9: 99
- 1849 Yuan, M., Ngou, B.P.M., Ding, P., and Xin, X.-F. (2021). PTI-ETI crosstalk: an integrative view of plant
1850 immunity. *Current Opinion in Plant Biology* 62, 102030. 10.1016/j.pbi.2021.102030.
- 1851 Yue et al. 2012. Widespread impact of horizontal gene transfer on plant colonization of land. *Nature*
1852 *Communications* 3:1152
- 1853 Yue et al. 2014. Origin of plant auxin biosynthesis. *Trends in Plant Science.* 19:764-770
- 1854 Żabka A, Polit JT, Winnicki K, Paciorek P, Juszczak J, Nowak M, Maszewski J. 2016. PIN2-like proteins
1855 may contribute to the regulation of morphogenetic processes during spermatogenesis in *Chara*
1856 *vulgaris*. *Plant Cell Reports* 35, 1655–1669.
- 1857 Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., and Yin, Y. (2018).
1858 dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids*
1859 *Research* 46, W95–W101. 10.1093/nar/gky418.
- 1860 Zhang, M., Zhang, Y., Scheuring, C.F., Wu, C.-C., Dong, J.J., Zhang H.-B. (2012) Preparation of
1861 megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics
1862 research. *Nat Protoc.* 7(3):467-78. doi: 10.1038/nprot.2011.455.
- 1863 Zhao, L.-H., Zhou, X.E., Wu, Z.-S., Yi, W., Xu, Y., Li, S., Xu, T.-H., Liu, Y., Chen, R.-Z., Kovach, A., et
1864 al. (2013). Crystal structures of two phytohormone signal-transducing α/β hydrolases: karrikin-
1865 signaling KAI2 and strigolactone-signaling DWARF14. *Cell Res* 23, 436–439. 10.1038/cr.2013.19.
1866

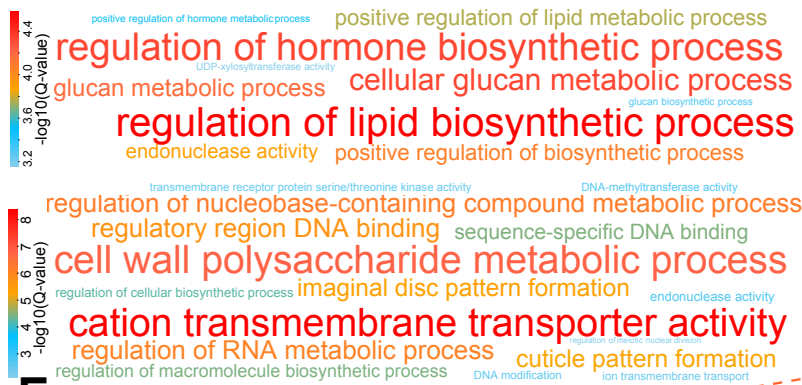




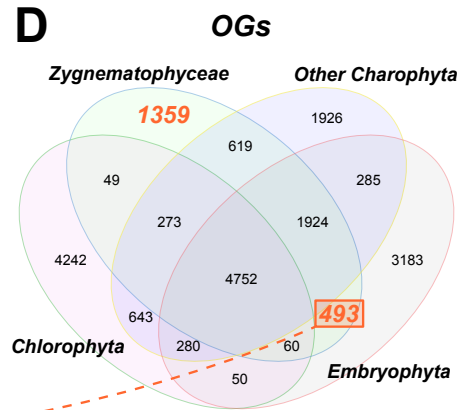
B Enriched Pfam domains (Z & E expanded OGs)



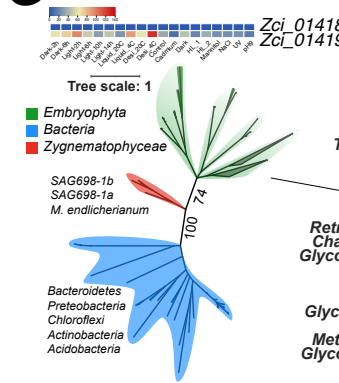
C Enriched GO functions (Z & E expanded OGs)



E Enriched GO functions (Z & E unique OGs)

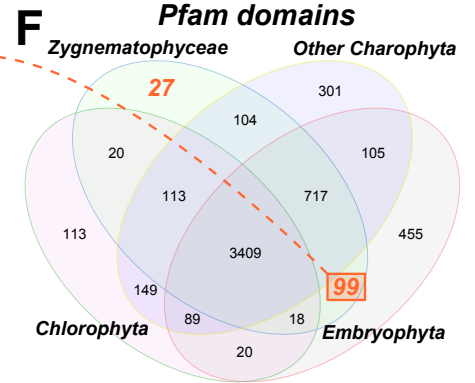


G Wound-induced protein WI12



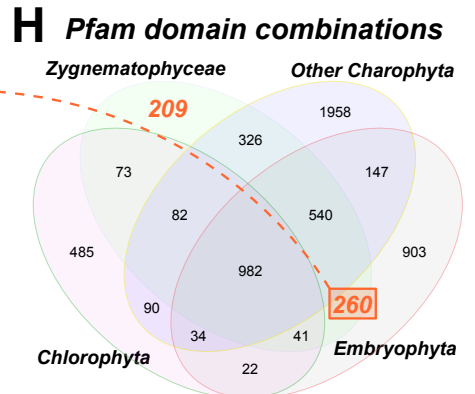
Selected Pfam domains present only in Z & E

	Pma	SAG698-1a	UTEX1559	Mpo	Ath
Inhibitor_I9	4	42	10	17	6
fn3_6	4	66	17	32	6
LOR	0	7	1	9	3
GRAS	23	291	9	14	11
DUF1218	3	3	1	2	1
FPP	3	0	1	2	2
DUF1685	5	2	1	1	1
Tmemb_185A	3	2	1	1	1
Ndr	3	10	1	1	2
DUF630	0	3	0	2	1
DUF632	0	3	0	2	1
WI12	0	0	1	2	2
SURNod19	0	0	1	5	3
Retrotran_gag_3	0	0	0	3	20
Chal sti synt N	5	18	2	2	2
Glyco_hydr_65N2	3	0	0	1	2
SMI1_KNR4	2	0	0	4	1
CRR7	3	3	0	1	1
Wound_ind	0	1	1	1	0
Glyco_hydr_30_2	0	12	1	4	3
MASE1	2	0	1	1	2
Methyltransf_1N	3	0	1	1	1
Glyco_hydro_3_C	0	1	1	1	1
Smu	1	1	2	1	1
Men					
SAG698-1b					
UTEX1560					
Ppa					

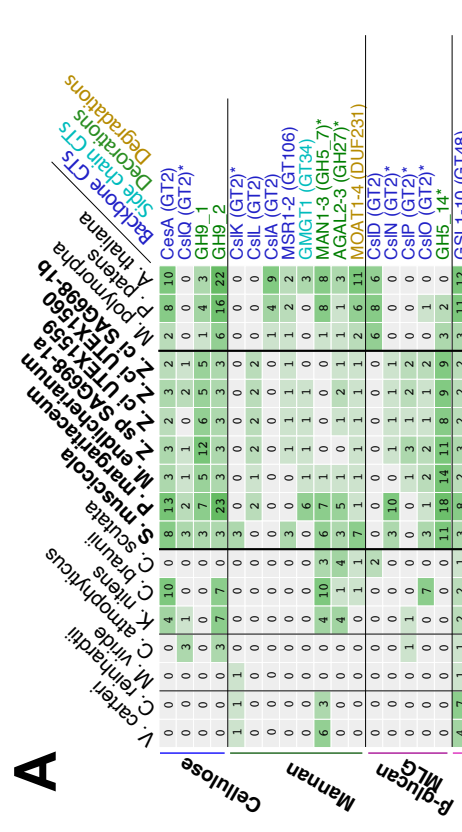
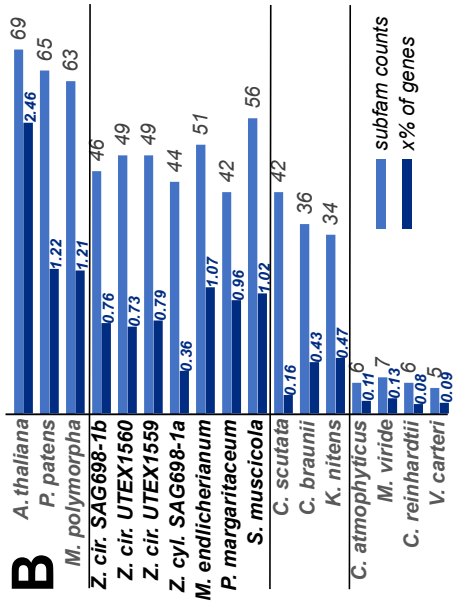
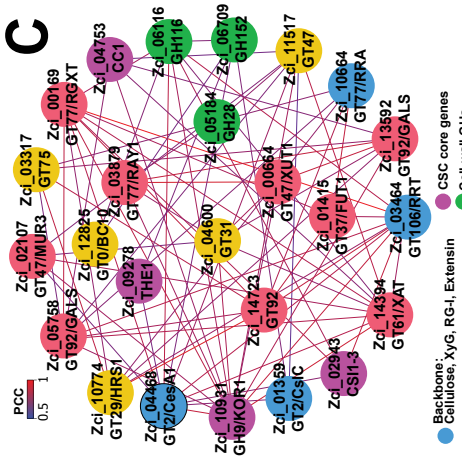
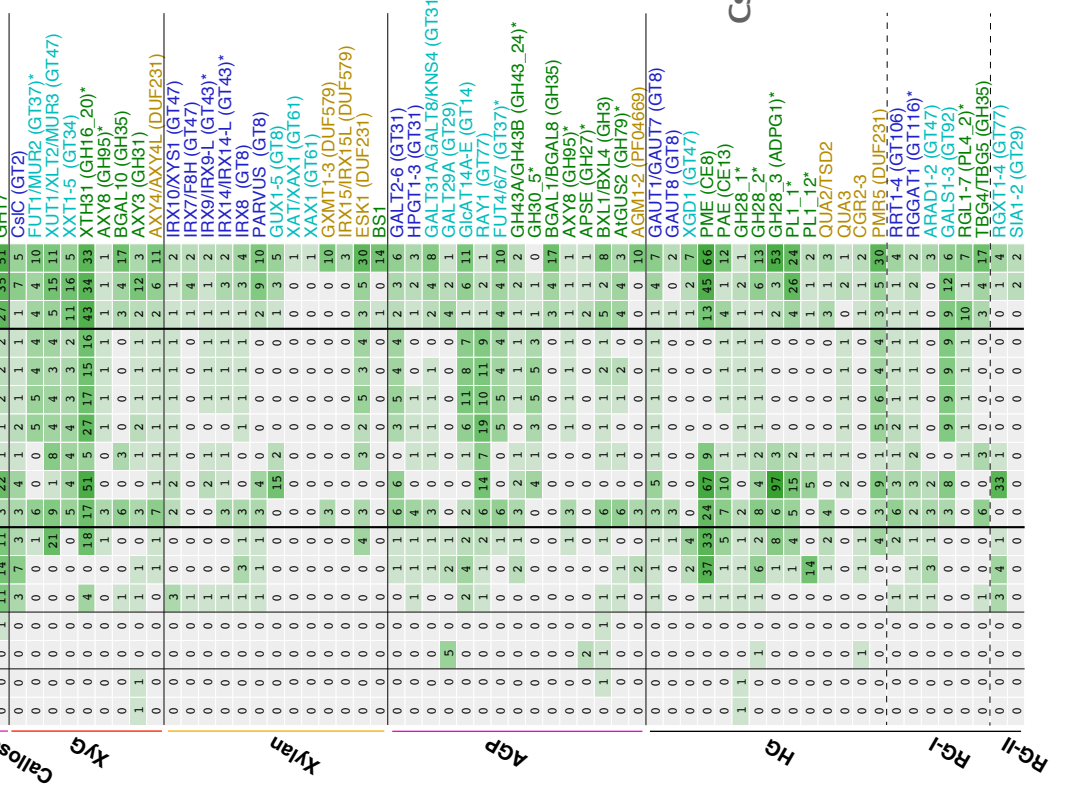
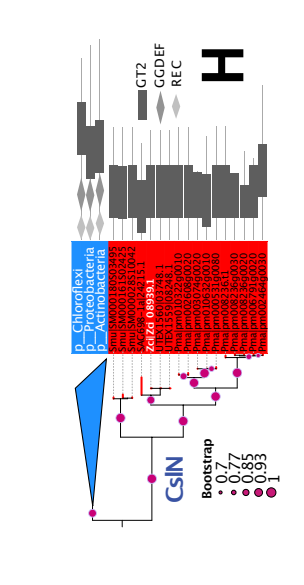
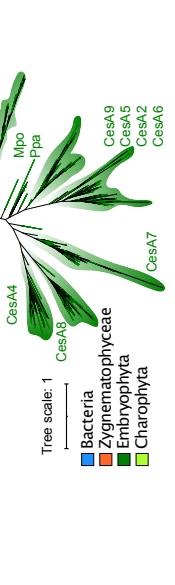
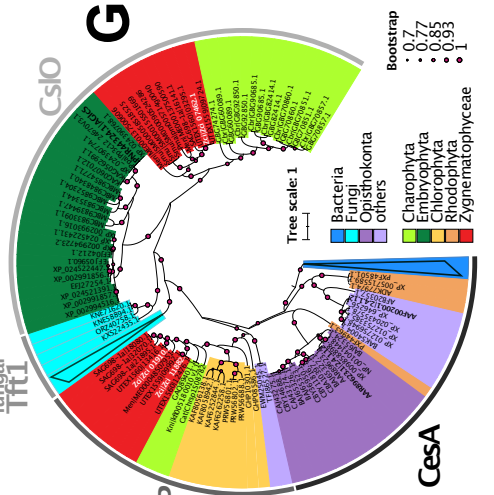
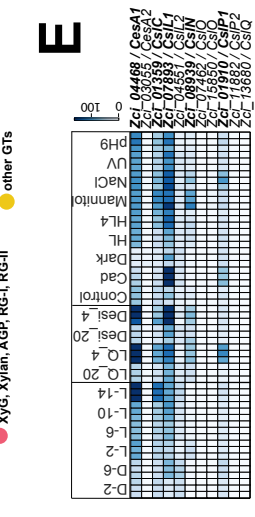
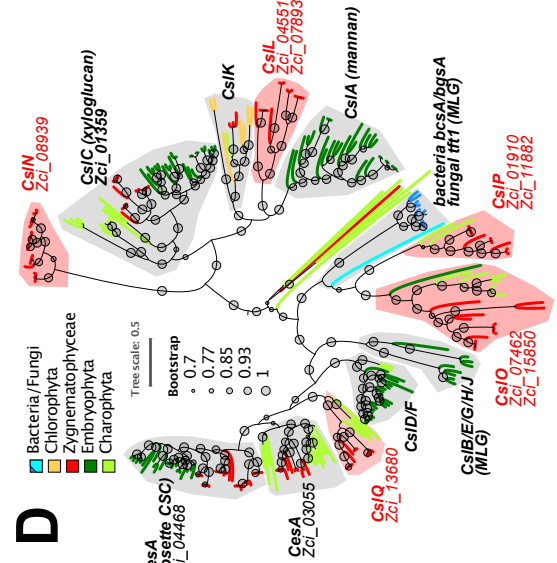
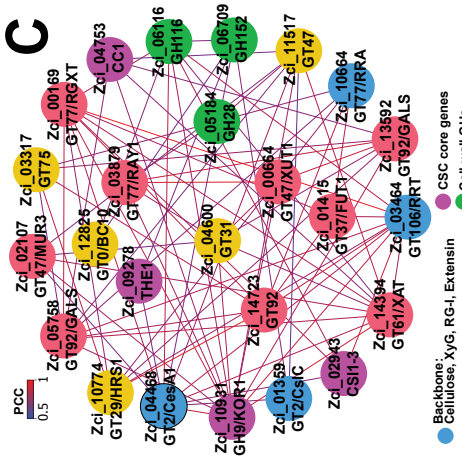
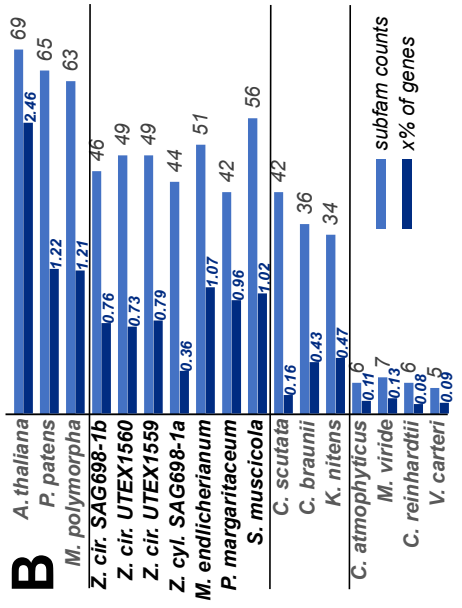


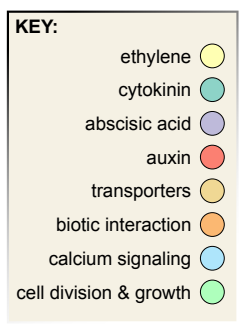
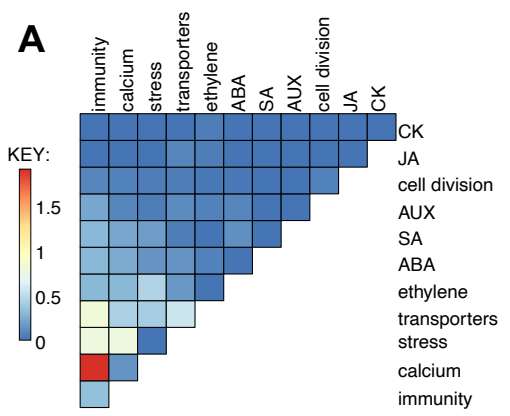
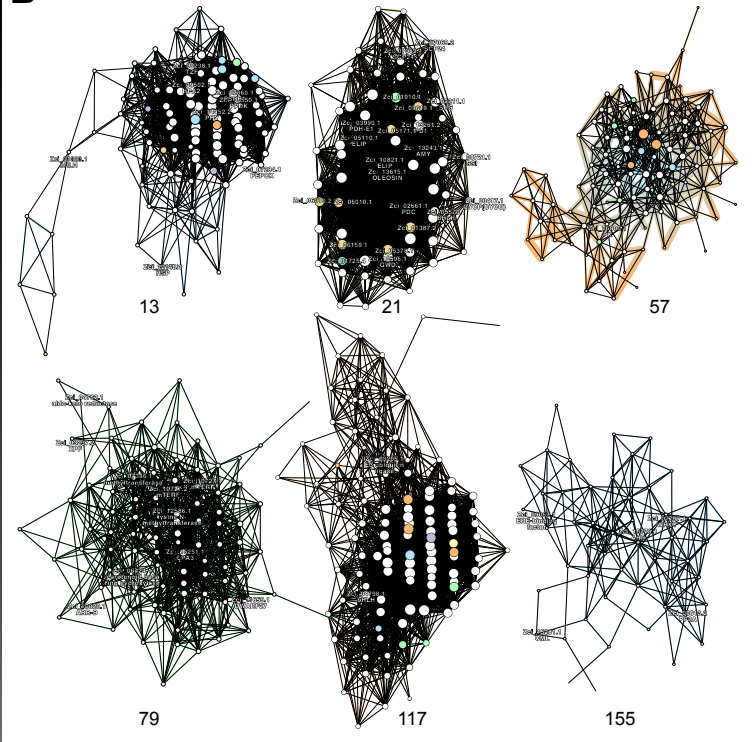
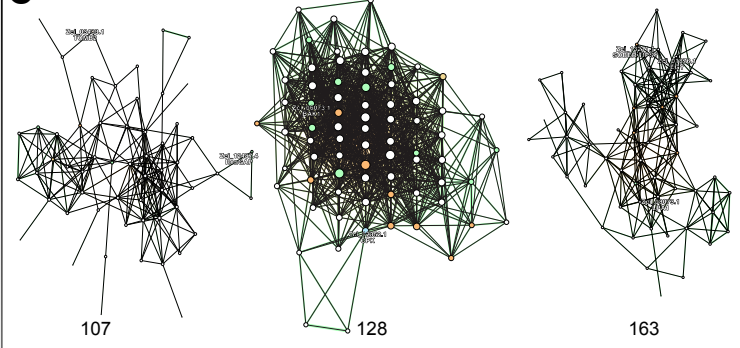
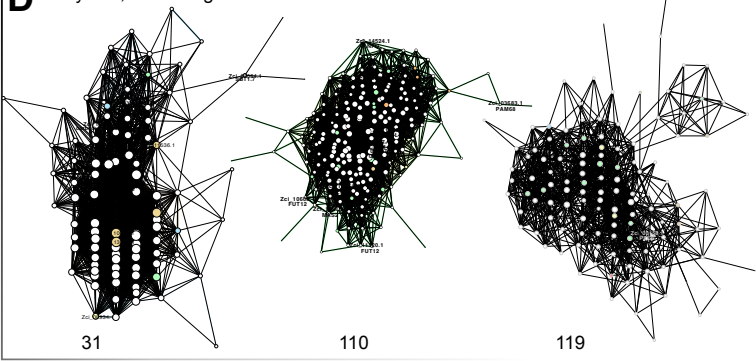
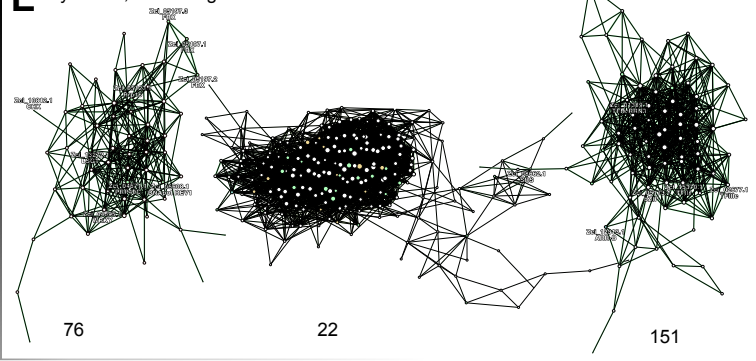
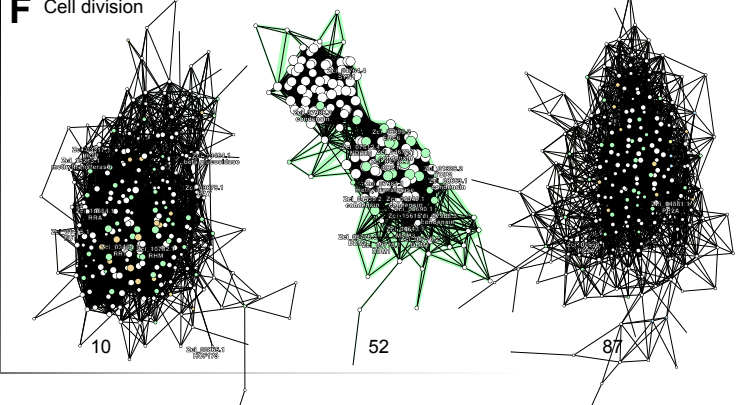
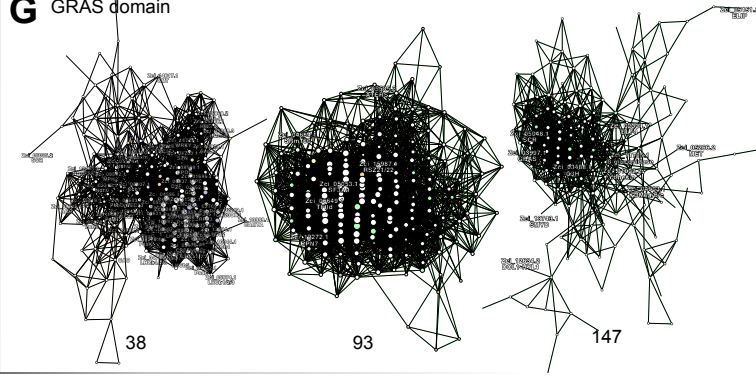
Selected Pfam domain combinations present only in Z & E

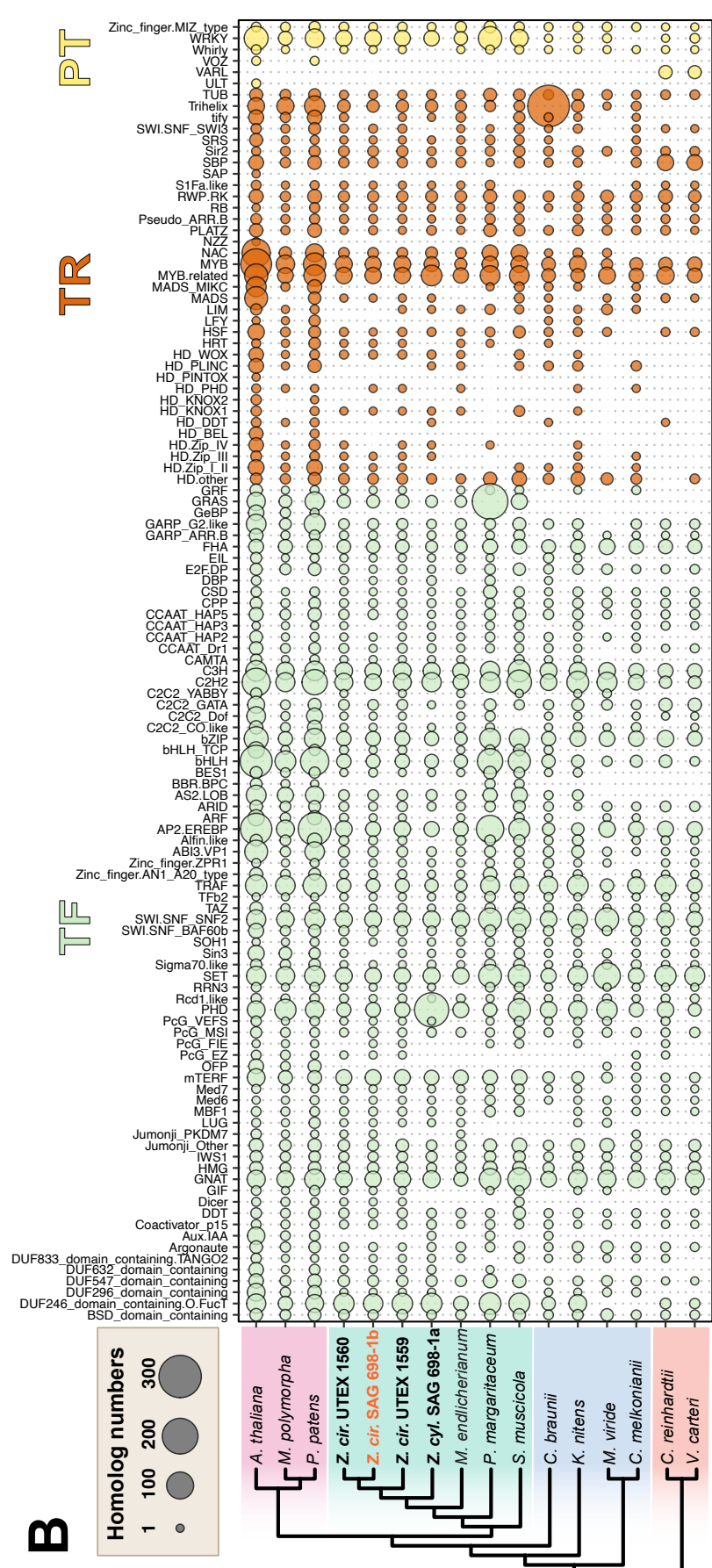
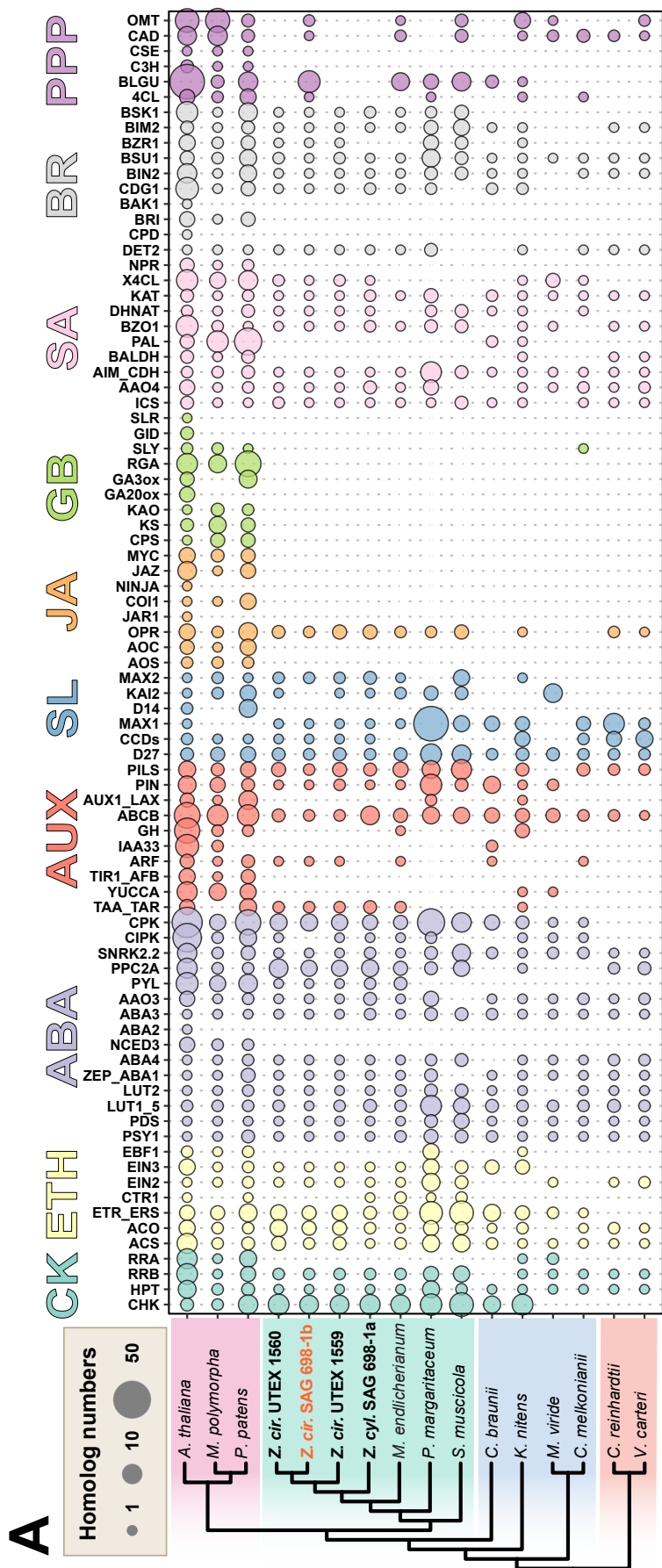
	Pma	SAG698-1a	UTEX1559	Mpo	Ath
SPOC;TFIIS_M	3	2	1	1	1
DUF3419;Ubie_methyltran	1	3	1	1	1
Sec23_helical;Sec23_trunk;zf-Sec23_Sec24	0	1	0	1	1
Tmemb_185A;zf-C3HC4_3	2	1	1	0	1
LRRNT_2;LRR_1;LRR_8;PKinase	1	1	0	6	6
Inhibitor_I9;PA;Peptidase_S8;fn3_6	12	13	4	11	2
LRRNT_2;LRR_1;LRR_8	2	6	1	13	12
Lectin_legB;PKinase	0	0	0	1	1
Chal sti synt C;Chal sti synt N	2	12	1	1	2
Cellulase;Glyco_hydro_5_C	0	1	1	2	1
Inhibitor_I9;Peptidase_S8	0	7	1	1	2
Inhibitor_I9;Peptidase_S8;fn3_6	2	12	1	4	1
LRRNT_2;LRR_1;LRR_8;PK_Tyr_Ser-Thr	2	0	0	2	1
Glyco_trans_4_5;Glycos_transf_1	3	3	0	1	2
PK_Tyr_Ser-Thr;U-box	3	3	0	1	2
Smu	3	3	0	1	2
Men					
SAG698-1b					
UTEX1560					
Ppa					

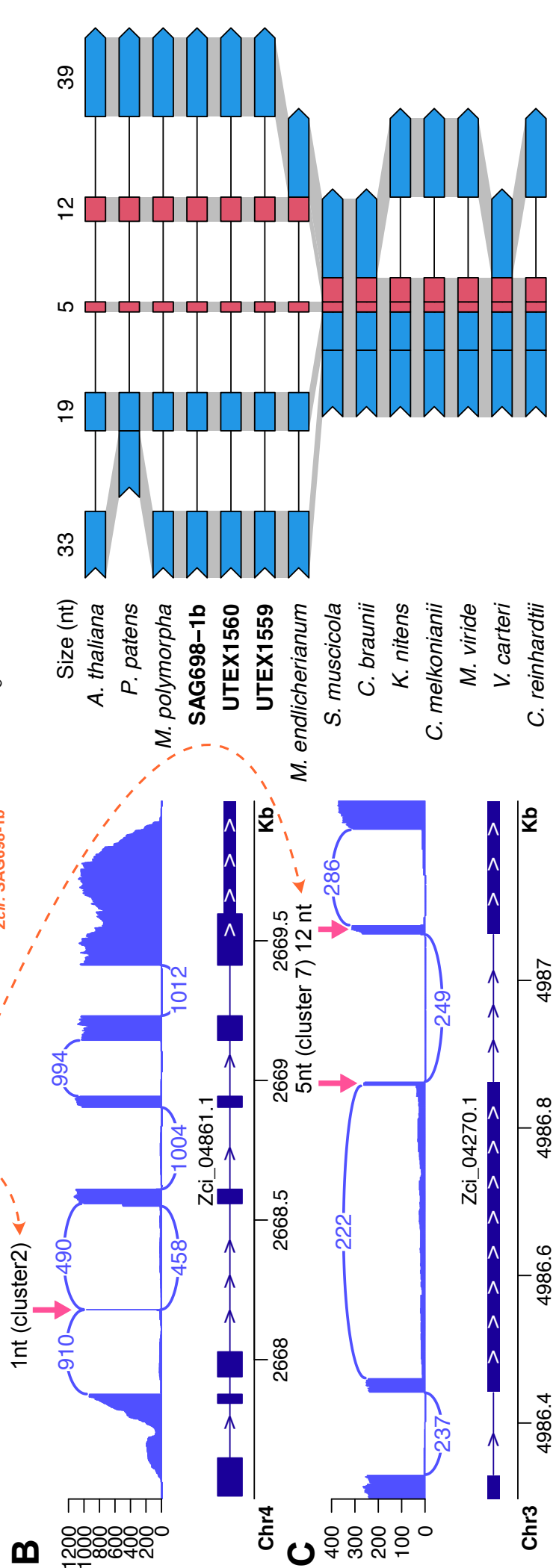
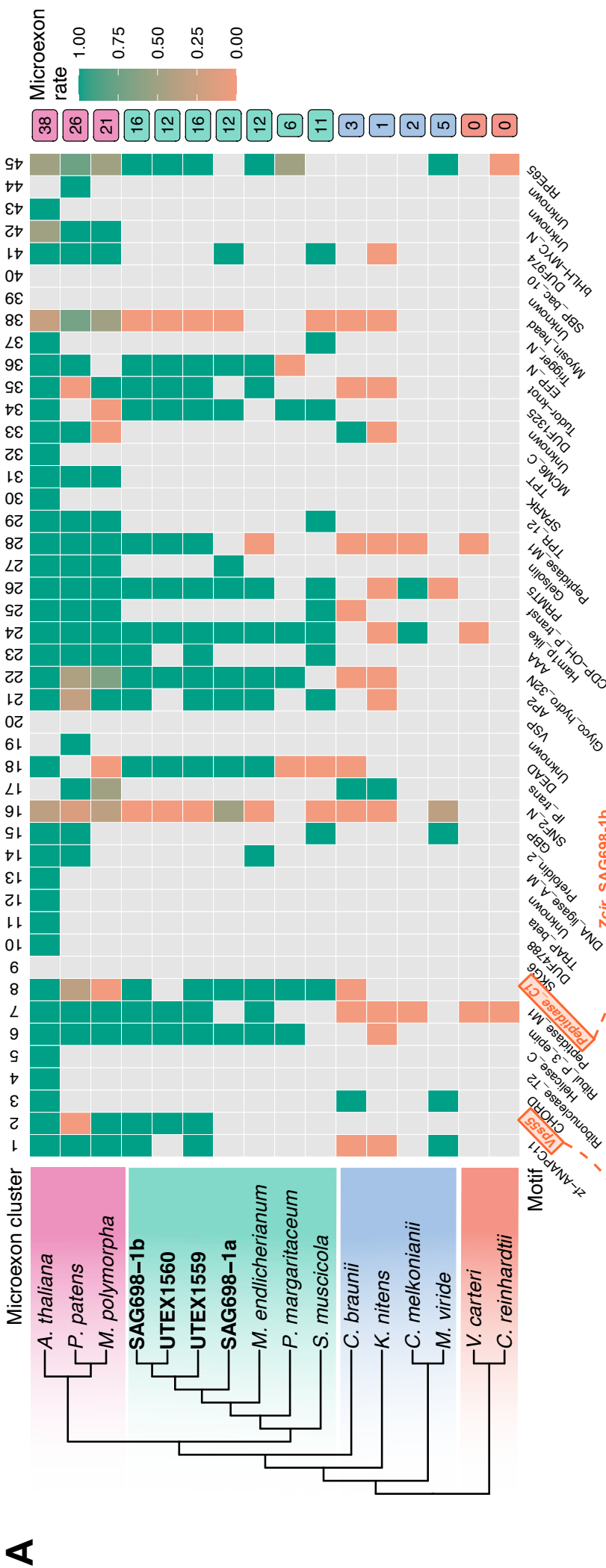


	<i>V. carteri</i>	<i>C. reinhardtii</i>	<i>C. atropathicus</i>	<i>K. nitens</i>	<i>C. braunii</i>	<i>C. scutata</i>	<i>S. muscicola</i>	<i>P. margharitaraneum</i>	<i>M. endlicherianum</i>	<i>Z. cyl. SAG698-1a</i>	<i>Z. cir. UTEX1559</i>	<i>Z. cir. UTEX1560</i>	<i>M. polymorpha</i>	<i>P. patens</i>	<i>A. thaliana</i>
Cellulose	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mannan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P-glucan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MLG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Callose	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XYG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Xylan	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AGP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RG-I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RG-II	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



A**B** Connectivity among signaling pathways**C** Biotic interaction**D** Ethylene, stress & growth**E** Cytokinin, stress & growth**F** Cell division**G** GRAS domain





4.5 **Publication VIII:** Unexpected cryptic species among streptophyte algae most distant to land plants

This research paper was published in the Journal “Proceedings of the Royal Society B: Biological Sciences” in November 2021. The full article as well as all supplementary figures and supplementary datasets can be found online:

<https://doi.org/10.1098/rspb.2021.2168>

Contribution of Janine Fürst-Jansen, middle author

J. M. R. Fürst-Jansen cultured the algal strains in the focus of this publication prior to RNA-Isolation. She then established a protocol for RNA extraction for *Chlorokybus* spp., carried out the extraction of the total RNA from all ten *Chlorokybus* isolates and quality checked the RNA prior to sending the samples to the sequencing facility. J. M. R. Fürst-Jansen critically read and revised the manuscript.



Cite this article: Irisarri I, Darienko T, Pröschold T, Fürst-Jansen JMR, Jamy M, de Vries J. 2021 Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc. R. Soc. B* **288**: 20212168. <https://doi.org/10.1098/rspb.2021.2168>

Received: 30 September 2021

Accepted: 1 November 2021

Subject Category:

Evolution

Subject Areas:

evolution, genomics, plant science

Keywords:

plant terrestrialization, plant evolution, streptophytes, green algae, phylogenomics, phycology

Authors for correspondence:

Iker Irisarri

e-mail: irisarri.iker@gmail.com

Tatyana Darienko

e-mail: tdarien@gwdg.de

Jan de Vries

e-mail: devries.jan@uni-goettingen.de

[†]These authors contributed equally to this study.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5705241>.

Unexpected cryptic species among streptophyte algae most distant to land plants

Iker Irisarri^{1,2,†}, Tatyana Darienko^{1,5,†}, Thomas Pröschold⁴,
Janine M. R. Fürst-Jansen¹, Mahwash Jamy⁶ and Jan de Vries^{1,2,3}

¹Department of Applied Bioinformatics, Institute for Microbiology and Genetics, ²Campus Institute Data Science (CIDAS), and ³Göttingen Center for Molecular Biosciences (GZMB), Department of Applied Bioinformatics, University of Goettingen, Goldschmidstrasse 1, 37077 Göttingen, Germany

⁴Research Department for Limnology, Leopold-Franzens-University of Innsbruck, Mondseestrasse 9, 5310 Mondsee, Austria

⁵Albrecht-von-Haller-Institute of Plant Sciences, Experimental Phycology and Culture Collection of Algae, University of Goettingen, Nikolausberger Weg 18, 37073 Göttingen, Germany

⁶Department of Organismal Biology, Program in Systematic Biology, Uppsala University, Norbyvägen 18D, 75236 Uppsala, Sweden

II, 0000-0002-3628-1137; TD, 0000-0002-1957-0076; TP, 0000-0002-7858-0434; JMR-F-J, 0000-0002-5269-8725; MJ, 0000-0002-2930-9226; JdV, 0000-0003-3507-5195

Streptophytes are one of the major groups of the green lineage (Chloroplastida or Viridiplantae). During one billion years of evolution, streptophytes have radiated into an astounding diversity of uni- and multicellular green algae as well as land plants. Most divergent from land plants is a clade formed by Mesostigmatophyceae, *Spirotaenia* spp. and Chlorokybophyceae. All three lineages are species-poor and the Chlorokybophyceae consist of a single described species, *Chlorokybus atmophyticus*. In this study, we used phylogenomic analyses to shed light into the diversity within *Chlorokybus* using a sampling of isolates across its known distribution. We uncovered a consistent deep genetic structure within the *Chlorokybus* isolates, which prompted us to formally extend the Chlorokybophyceae by describing four new species. Gene expression differences among *Chlorokybus* species suggest certain constitutive variability that might influence their response to environmental factors. Failure to account for this diversity can hamper comparative genomic studies aiming to understand the evolution of stress response across streptophytes. Our data highlight that future studies on the evolution of plant form and function can tap into an unknown diversity at key deep branches of the streptophytes.

1. Background

Green algae and land plants (Chloroplastida or Viridiplantae) consist of three major lineages: the recently pinpointed Prasinodermophyta [1], Chlorophyta and Streptophyta [2]. Streptophyta are about a billion years old [3,4] and encompass the main constituents of the land flora, the Embryophyta (land plants). In addition, Streptophyta include the algal relatives of land plants, known as streptophyte algae. In the past few years, the phylogenetic backbone of the green lineage has been brushed up. This was largely thanks to both an increased effort in sequencing streptophyte algae [5–13] and the use of these data in phylogenomic analyses to infer a robust green tree of life [2,14,15]. The new phylogenetic framework marked a milestone; it clarified the phylogenetic relationships among land plants and their streptophyte algal relatives. Within streptophytes, the position of Zygnematomophyceae as closest relatives to land plants made quite a splash. However, equally important was the

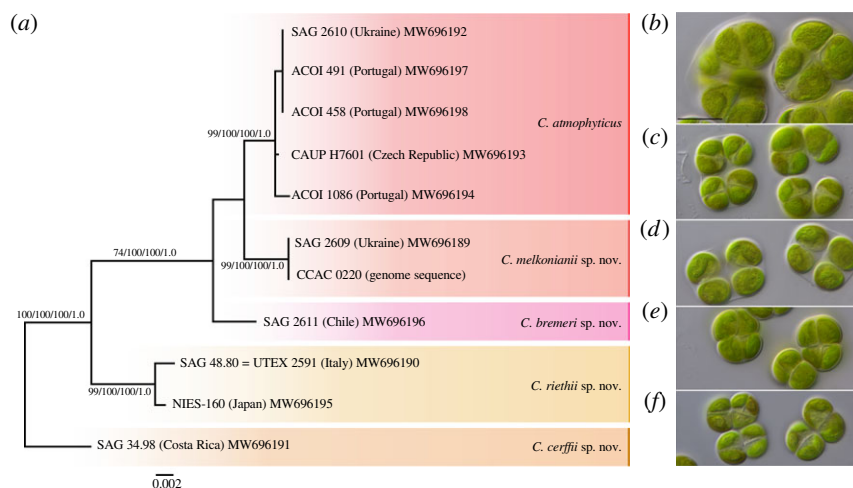


Figure 1. Cryptic diversity in *Chlorokybus*. (a) Maximum-likelihood phylogeny of SSU + ITS rDNA from all eleven isolates currently available in culture (root *sensu* figure 3). Branch support are respectively non-parametric bootstrap proportions from neighbour-joining, maximum parsimony, maximum likelihood and Bayesian posterior probabilities and branch lengths are in expected substitutions per site. Light micrographs correspond to: (b) *C. atmophyticus* ACOI 1086, (c) *C. melkonianii* sp. nov. SAG 2609, (d) *C. bremeri* sp. nov. SAG 2611, (e) *C. riethii* sp. nov. NIES-160, (f) *C. cerffii* SAG 34.98. Scale bar = 10 µm. (Online version in colour.)

recovery of Mesostigmatophyceae, *Spirotaenia* spp. [2] and Chlorokybophyceae as sister to all other Streptophyta [16]. Both Chlorokybophyceae and Mesostigmatophyceae are thought to encompass, respectively, one or few extant species. The apparent low diversity in these key lineages complicates macroevolutionary studies that aim to reconstruct the early evolution of key traits in the streptophyte ancestor. Recent genomic and phylogenomic investigations have honed in on freshwater and terrestrial streptophyte algae because they provide important insights into the origin of land plants and the evolution of response mechanisms to terrestrial stressors [2,5,7,10,12,13,17].

Here, we investigate the diversity within the Chlorokybophyceae using a phylotranscriptomic approach with broad sampling of isolates across its known distribution (Eurasia, Central and South America). We pinpoint that the Chlorokybophyceae consist of a cryptic species complex of at least five extant members.

2. Results and discussion

(a) Chlorokybophyceae is an oligotypic class

Chlorokybophyceae is thought to be a monotypic class with a single described species, *Chlorokybus atmophyticus* Geitler 1942. *Chlorokybus* is a subaerial alga inhabiting soil and rock surfaces and cracks [18–21]; it has been isolated from Europe and Central and South America, although it is thought to have a cosmopolitan distribution, despite being rare (electronic supplementary material, 'Portrait and history of *Chlorokybus*'). To further explore the distribution and diversity of *Chlorokybus*, we searched four large soil environmental sequencing datasets (Neotropical forest, Swiss Alps, meadow and agricultural soils from the UK and Tibet, and a set of globally distributed soils; approximately 128 Mio. reads total). Only a single amplicon sequence variant (ASV) of *Chlorokybus* was obtained, which was composed of 32 reads total (less than 0.01% abundance; electronic supplementary material, table S1). This ASV originated from a high-altitude Swiss Alpine soil sample [22]. Phylogenetic analyses confirmed the identity of this ASV as *Chlorokybus*, but its precise phylogenetic position could not

be determined because the SSU V4 region has limited phylogenetic signal [23] (electronic supplementary material, figure S1). None of the primer sets used in the above studies were biased against *Chlorokybus* and DNA extraction methods are unlikely to be so, but the lack of rocky outcrop samples in the above studies could have exacerbated the reported low abundance. Currently, 11 strains of *Chlorokybus* are available in public culture collections, none of them were isolated from the type locality and therefore no authentic strain is available (electronic supplementary material, table S2). We performed a phylogenetic analysis including all available *Chlorokybus* strains with two commonly used nuclear markers (SSU and ITS rDNA). This phylogeny suggested a deep genetic structure within *Chlorokybus* (figure 1a). Extensive observations under light microscope revealed no obvious morphological differences among the studied isolates, despite marked genetic divergences: all studied *Chlorokybus* isolates form sarcinoid, cubical packets of two to eight cells with a gelatinous matrix; cells are spherical or broadly ellipsoidal and contain a parietal slightly lobated chloroplast with two types of pyrenoids (figures 1 and 2; electronic supplementary material, figure S3; full description is provided below). The life cycle is haploid and was studied by Rieth [21] (figure 2). Since the phenotype did not give away hints as to the differences among the *Chlorokybus* strains, we garnered more sequence data.

(b) A phylotranscriptomic framework for *Chlorokybus*

Using the Illumina NovaSeq6000 platform, we generated 224 million paired-end reads (greater than 47 Gbp of raw sequence information) on four isolates of *Chlorokybus* from across its known distribution range. Combining these data with published genomic and transcriptomic information from other algae and land plants (electronic supplementary material, table S3), we inferred a robust phylogenomic tree based on 529 densely sampled loci (17% missing data). The maximum-likelihood tree, which was inferred with IQ-TREE under the LG + F + I + Γ 4 + C60 mixture model, unambiguously recapitulated the accepted phylogeny of the green lineage (Chloroplastida), including the position of *Chlorokybus* (Chlorokybophyceae), *Mesostigma* (Mesostigmatophyceae) and *Spirotaenia minuta* as the sister group to all other

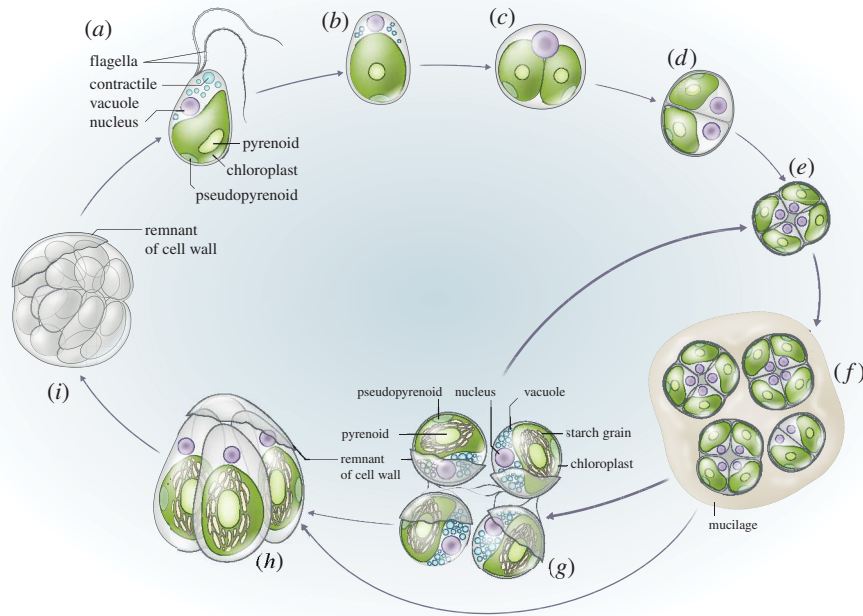


Figure 2. Life cycle of *Chlorokybus*. (a) Zoospore with two unilaterally inserted flagella in slightly under apical position. (b) A young vegetative cell is formed after the zoospore is settled and (c) cell division can begin. (d) Two-cell stage of daughter cells are contained within the same gelatinous matrix and (e) cubic cell packages can contain groups of two to eight cells each. (f) Mature packages produce mucilage and (g) cell cycle can proceed through the production of autospores for asexual reproduction (g to e). (h) Zoospores might be formed by differentiation from autospores (g to h) or directly from mature packages (f to h). (i) Zoospores can form groups of up to 32 cells called 'Maulbeerstadium'. Cell cycle based on Rieth [21]. (Online version in colour.)

streptophytes [1,13,14] (figure 3a). A summary coalescent analysis recovered an almost identical tree topology, except the interrelationships between SAG 34.98 (*C. cerffii* sp. nov.) and NIES-160 and UTEX 2591 (*C. riethii* sp. nov.) could not be resolved with certainty (electronic supplementary material, figure S2). Our phylotranscriptomic trees show unmistakable deep genetic structure within *Chlorokybus*, represented here by eight isolates. The genetic distances among *Chlorokybus* isolates are often more than twice as those recovered among three different species of *Arabidopsis* (figure 3a). The inferred patristic (maximum-likelihood) distances among *Chlorokybus* species are between 0.0254 and 0.0874 substitutions per site (*p*-uncorrected distances: 0.0245–0.0730), whereas the distances among the three *Arabidopsis* species are between 0.0149 and 0.0346 (*p*-uncorrected distances: 0.0147–0.0332) (table 1). A Bayesian relaxed molecular clock analysis calibrated with eight fossils (uniform priors) found that divergences within *Chlorokybus* could be as old as 76 Ma (95% HPD interval: 54–102 Ma) and the divergence between the two closest isolates described here as species—*C. atmophyticus* and *C. melkonianii* sp. nov., see below—was 24 Ma (95% HPD 15–34 Ma) (electronic supplementary material, figure S4). The use of more informative prior distributions for fossil calibrations (*t*-cauchy and skew-*t*) produced slightly younger divergences, as expected, but differences were not substantial (average differences within *Chlorokybus* were 0.47 and 1.47 Ma, respectively) (electronic supplementary material, figure S4). In contrast to *Chlorokybus*, the divergences among *Arabidopsis* species were 13 Ma (95% HPD 7–19 Ma) and 28 Ma (95% HPD 18–39 Ma).

To further scrutinize the deep genetic structure within *Chlorokybus*, we performed a maximum-likelihood analysis of 75 plastid proteins using IQ-TREE and the best-fit cpREV + F + I + Γ 4 + C60 mixture model. The plastid phylogeny

was moderately resolved and statistically supported; it further confirmed the deep divergences among *Chlorokybus* isolates, even though internal relationships in *Chlorokybus* differed from the nuclear tree (electronic supplementary material, figure S5). Similar plastid-nuclear incongruences are often observed in algae, for example in Volvocales [24], and might be due to either methodological or biological reasons (e.g. introgression), or both. While biological confounding factors cannot be excluded, the failure to recover *Amborella* as sister to all other flowering plants suggests the presence of biases and/or limited phylogenetic signal in the plastid dataset. At any rate, both plastid and nuclear marker phylogenies agreed on the presence of deep divergences among *Chlorokybus* isolates.

The final assessment of the genetic diversity within *Chlorokybus* is based on the more robust nuclear phylotranscriptomic dataset. On the basis of the inferred deep divergences, we propose a new taxonomic arrangement by describing four new species and assigning a lectotype and an epitype for *C. atmophyticus*, for which no authentic strain is available in public culture collections (see 'Systematic botany').

Taking advantage of the fact that the new isolates were grown simultaneously under the same experimental conditions, we explored whether the genetic distances among species are reflected in differences in global gene expression patterns. Clean Illumina reads were mapped against the annotated *Chlorokybus* genome using STAR [25] and expression quantified with RSEM [26], followed by TMM (trimmed mean of *M*-values) cross-sample normalization. While the lack of biological replicates prevented us from inferring differential gene expression, we observed marked differences in steady-state gene expression levels among the four new isolates (figure 3b, c). The clustering of expression values mirrored the species phylogeny, with NIES-160 (*C. riethii* sp. nov.) showing the most

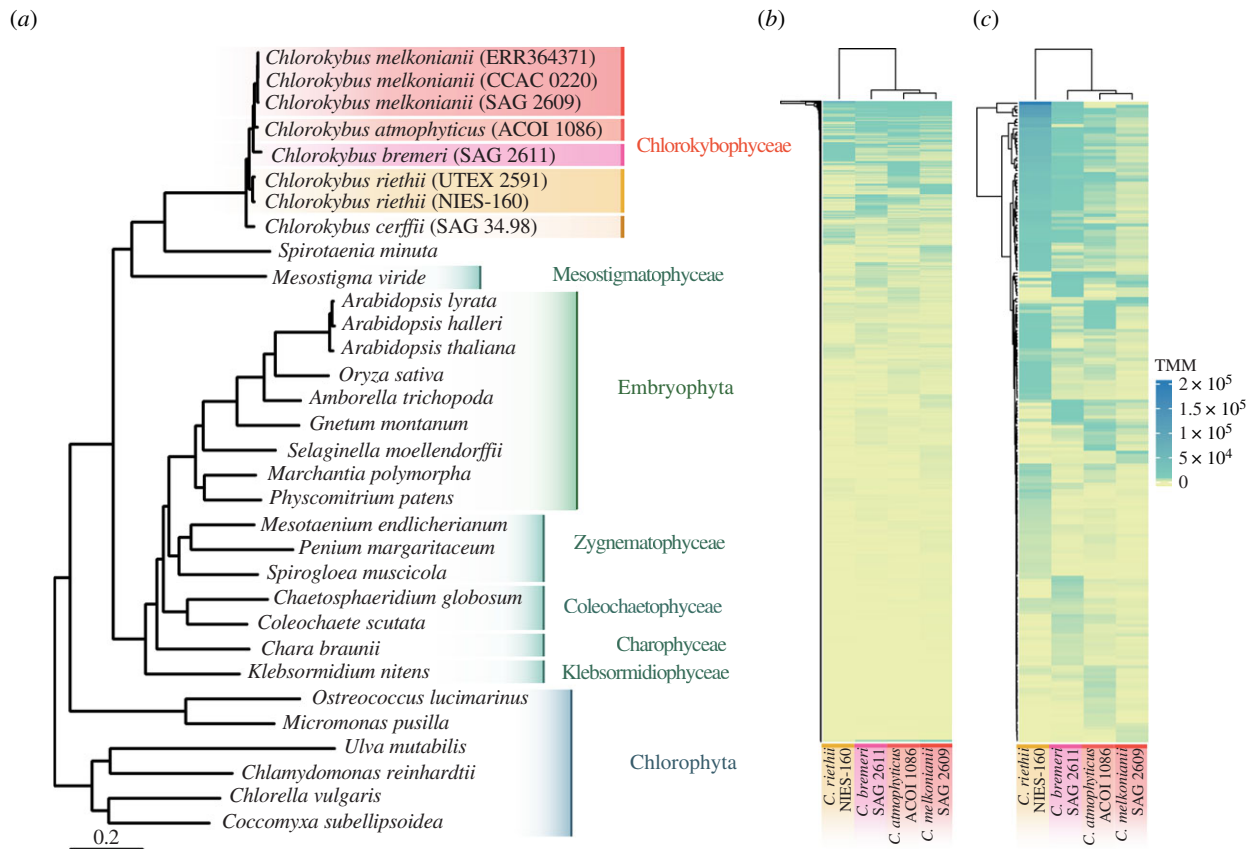


Figure 3. Transcriptomic evidence for deep phylogenetic divergences and expression differences within *Chlorokybus*. (a) Maximum-likelihood phylogeny based on 529 densely sampled loci, inferred with IQ-TREE under LG + F + I + Γ 4 + C60 and support values from 1000 pseudoreplicates of UFBoot2 and SH-aLRT (all branches received 100% support). Branch lengths are in expected substitutions per site. (b,c) Gene expression differences (TMM, trimmed mean of M -values) among four isolates grown simultaneously under the same experimental conditions. Heatmaps correspond to (b) the 9300 annotated proteins in the *C. melkonianii* genome (no filtering) and (c) the top 200 proteins with the highest expression differences. (Online version in colour.)

Table 1. Genetic distances among *Chlorokybus* isolates and *Arabidopsis* species measured from concatenated amino acid alignments of 529 loci (178 397 aligned amino acids). p -uncorrected (upper triangle) and maximum-likelihood distances (lower triangle; figure 1) are shown, with intra-specific comparisons in italics.

	1	2	3	4	5	6	7	8
1	<i>C. cerffii</i> (SAG 34.98)	0.0573	0.0619	0.0781	0.0690	0.0718	0.0730	0.0722
2	<i>C. riethii</i> (NIES-160)	0.0621	<i>0.0133</i>	0.0646	0.0507	0.0482	0.0539	0.0522
3	<i>C. riethii</i> (UTEX 2591)	0.0677	<i>0.0135</i>	0.0648	0.0501	0.0501	0.0529	0.0517
4	<i>C. bremeri</i> (SAG 2611)	0.0874	0.0710	0.0713	0.0424	0.0446	0.0497	0.0486
5	<i>C. atmophyticus</i> (ACOI 1086)	0.0762	0.0543	0.0536	0.0452	0.0245	0.0289	0.0277
6	<i>C. melkonianii</i> (ERR364371)	0.0798	0.0514	0.0536	0.0479	0.0254	<i>0.0044</i>	<i>0.0002</i>
7	<i>C. melkonianii</i> (SAG 2609)	0.0811	0.0580	0.0569	0.0537	0.0302	<i>0.0045</i>	<i>0.0037</i>
8	<i>C. melkonianii</i> (CCAC 0220)	0.0801	0.0559	0.0553	0.0523	0.0288	<i>0.0002</i>	<i>0.0037</i>
	1	2	3					
1	<i>A. halleri</i>	0.0332	0.0147					
2	<i>A. thaliana</i>	0.0346	0.0288					
3	<i>A. lyrata</i>	0.0149	0.0296					

different expression profile, followed by SAG 2611 (*C. bremeri* sp. nov.), and the more similar profiles shown by ACOI 1086 (*C. atmophyticus*) and SAG 2609 (*C. melkonianii* sp. nov.). Yet, even the two latter isolates showed marked differences in gene expression, which together with the reported genetic distances support the notion that they are not only different species but might also exhibit different cell physiologies.

3. Conclusion

Here, we report on the presence of consistent deep structure within *Chlorokybus* after analysing all currently available isolates. These divergences might date back to approximately 76 Ma and are twice as large as those among some flowering plant species (e.g. *Arabidopsis*). Deep genetic divergences

among *Chlorokybus* isolates are further supported by substantial gene expression variation when grown under the same experimental conditions. Yet, these genetic differences are not reflected in appreciable morphological differences, which suggest the presence of undescribed cryptic diversity within this lineage. All this genetic diversity has remained unnoticed under the umbrella name *Chlorokybus atmophyticus*, the only validly described species so far. To remedy this, we describe four new species of *Chlorokybus* and designate a cryopreserved culture as epitype for *C. atmophyticus*. *Chlorokybus* species are probably cosmopolitan but rare, as further supported by our search across global soil metabarcoding datasets that identified a single sequence of this genus. Properly recognizing the existing diversity within *Chlorokybus* is paramount, given the key phylogenetic position of Chlorokybophyceae, which together with *Spirotaenia* spp. [27] and Mesostigmatophyceae are the sister lineage to all other streptophytes. This diversity has to be taken into account for the adequate comparison of current and future data from different *Chlorokybus* strains [2,8,13]. In fact, the reported gene expression differences might even suggest certain interspecific variability in responding to environmental factors and adequately accounting for this will be essential in comparative genomic studies that aim to understand the evolution of key traits (such as phytohormone or stress response pathways [17]) along the backbone phylogeny of streptophytes. Our phylogenetic analysis of genomic data can aid in uncovering key cryptic diversity, which together with the discovery of new deep-branching lineages [28–30], are revealing important pieces in the puzzle that is the Eukaryotic Tree of Life.

4. Systematic botany

In the following, we describe four new species of *Chlorokybus* and designate a lectotype and an epitype for *C. atmophyticus*, given that no cultured material is available from the different locations studied by Geitler [18–20]. We further provide a formal description of the class Chlorokybophyceae, which was originally proposed by Bremer [31] without formal description nor page numbers, and thus being invalid under articles 38.1 and 41.5 of the International Code of Nomenclature (ICN) for algae, fungi and plants [32].

Class Chlorokybophyceae class. nov. (figure 2)

Description: Green algae forming sarcinoid, cubical packets. Single chloroplast containing two pyrenoids. First pyrenoid located in the middle of the chloroplast and surrounded by starch grains. Second naked pyrenoid (or called pseudopyrenoid) located at the edge of the chloroplast. Reproduction can occur asexually by breaking cell packages into separate cells or by zoospores (figure 2). Zoospores are produced one per cell and possess two laterally inserted flagella. The flagella and body are covered with square scales. The flagellar apparatus is non-cruciate unilateral and contains multi-layered structures (MLS). After settling of the zoospores, the flagella are retracted at the point of their insertion. Cell division type phragmoplast-like, presence of advanced cleavage furrow and VII type of mitosis (*sensu* van den Hoek *et al.* [33]). Sexual reproduction is not observed. The class is supported by SSU rDNA, plastid and nuclear transcriptomic data.

Type order (designated here): Chlorokybales ordo nov.

Order Chlorokybales ordo nov.

Description: With features of the class.

Type family (designated here): Chlorokybophyceae fam. nov.

Family Chlorokybaceae fam. nov.

Description: With features of the class.

Type genus (designated here): *Chlorokybus* Geitler 1942, Österr. Bot. Z. 91: 51.

Comment: Rogers *et al.* (1980) proposed the order Chlorokybales and the family Chlorokybaceae without Latin diagnosis. They referred to the Latin diagnosis of Geitler (1942/43), but he published the Latin diagnosis in 1942 (see detailed citation below).

Type species: *Chlorokybus atmophyticus* Geitler 1942, Österr. Bot. Z. 91: 51; Geitler 1942/43, Flora 136, fig. 2 (lectotype designated here).

Emended description: Cell size 16.9–20.0 µm length × 12.0–16.5 µm wide. Other features are identical to the class description. SSU-ITS sequence (MW696194) and NCBI BioSample accession SAMN18221336 (RNA-Seq), ITS-2 Barcode: BC-1 in electronic supplementary material, figure S6.

Diagnosis: Differs from other species of *Chlorokybus* by SSU-ITS and transcriptome sequence.

Epitype (designated here): Strain ACOI 1086 cryopreserved in metabolically inactive state at the Culture Collection of Algae (SAG), Georg-August-University Göttingen, Germany (figure 1c; electronic supplementary material, figure S3a–d).

Chlorokybus melkonianii sp. nov.

Description: Cell size 10.3–13.5 µm length × 7.7–10.6 µm wide. Other features are identical to the class description. SSU-ITS sequence (MW696189), NCBI BioSample accession SAMN18221334 (RNAseq), ITS-2 Barcode: BC-2 in electronic supplementary material, figure S6.

Diagnosis: Differs from other species of *Chlorokybus* by SSU-ITS and transcriptome sequence.

Holotype (designated here): Strain SAG 2609 cryopreserved in metabolically inactive state at the Culture Collection of Algae (SAG), Georg-August-University Göttingen, Germany (figure 1b; electronic supplementary material, figure S3e–h).

Type locality: Europe, Ukraine, regional landscape park ‘Trakhtemyriv’, sandstone outcrops, in crack.

Etymology: The species epithet honours Prof. Dr Michael Melkonian (University of Cologne, Germany) for his important contributions to understanding the diversity and evolution of algae.

Comment: The strain CCAC 0220 represents another isolate of this species and the SSU-ITS sequence and NCBI BioSample accession are available under SAMEA2242428 (RNAseq) and SAMN10351691 (genome assembly), respectively.

Chlorokybus bremeri sp. nov.

Description: Cell size 13.1–16.8 µm length × 9.7–11.5 µm wide. Other features are identical to the class description. SSU-ITS sequence (MW696196) and NCBI BioSample accession SAMN18221335 (RNA-Seq), ITS-2 Barcode: BC-3 in electronic supplementary material, figure S6.

Diagnosis: Differs from other species of *Chlorokybus* by SSU-ITS and transcriptome sequence.

Holotype (designated here): Strain SAG 2611 cryopreserved in metabolically inactive state at the Culture Collection of Algae (SAG), Georg-August-University Göttingen, Germany (figure 1d; electronic supplementary material, figure S3i–l).

Type locality: South America, Chile, national park 'La Campana', granite outcrops, in crack.

Eymology: The species epithet honours Prof. Dr Kåre Bremer (University of Stockholm, Sweden), who first proposed the class name Chlorokybophyceae.

Chlorokybus riethii sp. nov.

Description: Cell size 13.1–16.8 µm length × 9.7–11.5 µm wide. Other features are identical to the class description. SSU-ITS sequence (MW696190) and NCBI accession SRX025846 (RNA-Seq), ITS-2 Barcode: BC-4a/b in electronic supplementary material, figure S6.

Diagnosis: Differs from other species of *Chlorokybus* by SSU-ITS and transcriptome sequence.

Holotype (designated here): Strain SAG 48.80 cryopreserved in metabolically inactive state at the Culture Collection of Algae (SAG), Georg-August-University Göttingen, Germany (figure 1e; electronic supplementary material, figure S3m–r).

Type locality: Europe, Italy, Naples, in enrichment culture of *Porphyridium purpureum* from the Mediterranean Sea.

Eymology: The species epithet honours Prof. Dr Alfred Rieth (Institute for Genetics and Crop Plant Research Gateleben) for his detailed observations of *Chlorokybus*.

Comment: The strain NIES-160 represents another isolate of this species and the SSU-ITS sequence and NCBI BioSample accession are available under MW696195 and SAMN18221337 (RNA-Seq), respectively.

Chlorokybus cerffii sp. nov.

Description: Cell size 13.1–16.8 µm length × 9.7–11.5 µm wide. Other features are identical to the class description. SSU-ITS sequence (MW696191) and NCBI BioSample accession SAMN07525888 (RNA-Seq), ITS-2 Barcode: BC-5 in electronic supplementary material, figure S6.

Diagnosis: Differs from other species of *Chlorokybus* by SSU-ITS and transcriptome sequence.

Holotype (designated here): Strain SAG 34.98 cryopreserved in metabolically inactive state at the Culture Collection of Algae (SAG), Georg-August-University Göttingen, Germany (figure 1f and electronic supplementary material, figure S3s–v).

Type locality: Central America, Costa Rica, Province Heredia, Barreal coffee plantation, soil.

Eymology: The species epithet honours Prof. Dr Rüdiger Cerff (Braunschweig University of Technology, Germany) for his contributions on endosymbiosis research and plant evolutionary biology.

5. Material and methods

(a) Culturing conditions

Details about isolate origins are available in electronic supplementary material, table S2. Four isolates (NIES-160, SAG 2611, ACOI 1086, SAG 2609) were cultivated on 3N-BBM + V medium (medium 26a in Schlösser [34]) at 18°C, with 20 µmol photons m⁻² s⁻¹ provided by daylight fluorescent tubes (TL-D 18 W 640, Osram, Munich, Germany), and light:dark cycle of 16:8 h. Data for the strain SAG 34.98 were obtained from de Vries *et al.* [8], cultured in ES (medium 1 in Schlösser [35]) at 20°C with 50 µmol of photons m⁻² s⁻¹ from LED light source and light:dark cycle of 12:12 h. Three-week-old cultures were used for morphological identification, comparing them to the original species descriptions. Light microscopy used an Olympus BX-60 microscope (Olympus, Tokyo, Japan), a ProgRes C14plus camera, and the ProgRes CapturePro imaging system (v2.9.0.1) (Jenoptik, Jena, Germany).

(b) rDNA phylogeny

DNA was extracted with the DNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The SSU and ITS were amplified using the *Taq* PCR MasterMix Kit (Qiagen) with primers EAF3 and ITS055R [36]. PCR reactions had initial denaturation at 95°C for 5 min followed by 30 cycles of 1 min at 95°C, 2 min at 55°C and 3 min at 68°C, and a final step at 68°C for 10 min. PCR products were purified and sequenced as in Darienko *et al.* [37]. A multiple sequence alignment of SSU was performed according to the predicted secondary structures (electronic supplementary material, figure S6). ITS-1 and ITS-2 were folded according to Darienko *et al.* [38]. SSU/ITS sequences were concatenated into a dataset containing 11 OTUs (2,424 bp). We used PAUP 4.0a build 169 [39] to select the best-fit evolutionary model (GTR + I) according to the Akaike information criterion (AICc). Neighbour-joining, maximum parsimony, maximum likelihood and Bayesian inference were conducted following Darienko *et al.* [34], using PAUP v4.0a build 169 [39], RAxML v8.2.12 [40], MrBayes v3.2.7a using the doublet approach [41] and PHASE package v2.0 [42].

(c) Metabarcoding

Environmental eukaryotic SSU amplicon sequences were obtained from previous studies [22,43–45] (electronic supplementary material, table S1). Short-read data were cleaned and denoised into Amplicon Sequence Variants (ASV) using DADA2 [46]. SSU of the *C. mekonianii* genome (RHPI01002076.1:1257–3060) was used to search the datasets with BLASTN v2.11.0+ [47] using a 95% sequence similarity threshold. Primers from the original studies were tested *in silico* with the TestPrime function in SILVA [48] to discard biases against *Chlorokybus*. The identified *Chlorokybus* ASV was aligned to other *Chlorokybus* rDNA using MAFFT v7.304b [49] (default settings) and a phylogeny was inferred with IQ-TREE v1.6.12 [50] using the BIC-selected model and 1000 non-parametric bootstrapping pseudoreplicates.

(d) RNAseq and transcriptome assembly

Algae were scraped off the agar and transferred into 1 ml Trizol (Thermo Fisher, Carlsbad, CA, USA), ground using a Tenbroek tissue homogenizer and RNA isolated according to the manufacturer's instructions. RNA samples were treated with DNase I (Thermo Fisher, Waltham, MA, USA) and quality and quantity assessed with a formamide agarose gel and Nanodrop (Thermo Fisher), respectively. RNA was shipped on dry ice to Genome Québec (Montreal, Canada). After Bioanalyzer (Agilent Technologies Inc., Santa Clara, CA, USA) quality check, libraries were built using the NEBNext mRNA stranded library preparation kit (New England Biolabs, Beverly, MA, USA). Libraries were checked with Bioanalyzer and sequenced using NovaSeq 6000 (Illumina) with NEBNext dual adapters: 5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3' for read 1 and 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3' for read 2. FastQC (www.bioinformatics.babraham.ac.uk/projects/) reports are available in Dryad.

We downloaded RNAseq data for *Chlorokybus atmophyticus* SAG 34.98 [8] (SRX3107749-SRX3107751), *Chlorokybus melkonianii* [2] (ERR364371), *Chaetosphaeridium globosum* [2] (ERR364369), and *Coleochaete orbicularis* [51] (SRR1594679). For all samples, transcriptomes were assembled *de novo* using Trinity v2.11.0 [52] after adapter trimming (—trimmomatic). SuperTranscripts [53] were inferred by collapsing splicing isoforms, as implemented in Trinity. Transcriptome completeness was assessed with BUSCO v4.1.0 [54] with the 'chlorophyta_odb10' reference set. All new assemblies recovered greater than 75% complete BUSCOs (electronic supplementary material, table S4). Protein-coding genes were identified with Transdecoder v5.5.0 using *Chlorokybus*'s

annotated proteins (CCAC 0220) as reference in BLASTP searches and retaining only the longest open reading frame (ORF) per transcript (--single_best_only). A total of 19 147 transcripts for *C. riethii* UTEX 2591 (published as *C. atmophyticus* [55]) were downloaded from GenBank, assembled from GS FLX Titanium 454.

(e) Phylotranscriptomic dataset construction

Likely contaminants were removed by sequence similarity searches against a database containing proteins from (i) *Chlorokybus melkonianii* (CCAC 0220) [13] and possible contaminants including (ii) RefSeq representative bacterial genomes (11 318 genomes) and (iii) fungi (2397), (iv) all available viruses (125), (downloaded from GenBank on 17/08/2020), (v) *Chlamydomonas reinhardtii*, and (vi) a *Vermamoeba vermiformis* transcriptome [56]. *Vermamoeba* and *Chlamydomonas* were identified as likely contaminants in some cultures. MMseqs2 [57] was used with an iterative search with increasing sensitivities, real sequence identity, and keeping 10 hits maximum (--start-sens 1 --sens-steps 3 -s 7 --alignment-mode 3 --max-seqs 10). As strict decontamination criterion, only sequences whose best hit corresponded to a predicted *Chlorokybus* nuclear proteins were kept for phylogenetic analyses (4817–19 566 proteins per species; 8690–10 917 for new transcriptomes).

(f) Phylotranscriptomic phylogeny

A representation of chlorophytes and streptophytes were used as outgroups (electronic supplementary material, table S3). Orthofinder v2.4.0 [58] was used to infer orthogroups using a species tree following Leebens-Mack *et al.* [2] with unresolved relationships within *Chlorokybus*. We selected phylogenetic hierarchical orthogroups at the tree root [58]. In total, 2386 orthogroups contained data for all major lineages (in practice, at least one sequence each for *Chlorokybus*, *Mesostigma* or *Spirotaenia minuta*, *Coleochaete* or *Chaetosphaeridium*, *Chara* or *Klebsormidium*, Zygnematophyceae, chlorophytes, bryophytes, and vascular plants). Homologous sets were aligned with MAFFT v7.304b [49] (default settings) and subjected to maximum-likelihood inference with IQ-TREE v1.6.12 [50] using fast searches, BIC-selected best-fit nuclear models, and SH-like aLRT branch support (-m TEST -msub nuclear -fast -alrt 1000). Phylopyrpruner v1.2.3 (Thalen *et al.*, <https://pypi.org/project/phylopyrpruner/>) was used to prune orthologue sets (--prune MI --mask pdist --trim-lb 5 --trim-freq-paralogues 4 --trim-divergent 1.25 --min-pdist 1×10^{-8} --min-support 0.75 --min-taxa 10 --min-gene-occupancy 0.1 --min-otu-occupancy 0.1), resulting in 946 orthologue sets. After applying the above taxonomic filter, we selected 529 final loci, which were masked with PREQUAL v1.02 [59], aligned with MAFFT ginsi v7.304b with a variable scoring matrix ('--allowshift --unalignlevel 0.8') [60], and columns greater than 75% gaps removed with ClipKIT v0.1 [61]. Trimmed alignments were concatenated into a matrix containing 32 taxa and 529 loci (17% missing sequences) and 178 397 aligned amino acid positions. Maximum-likelihood trees were inferred using IQ-TREE under BIC-selected homogeneous (LG + F + I + I4) and mixture (LG + F + I + I4 + C60) models and branch support assessed with 1000 pseudoreplicates of UFBoot2 [62] and SH-like aLRT [63]. ASTRAL v5.7.5 [64] was run on gene trees inferred by IQ-TREE with BIC-selected models (branches with less than 10% bootstrap were collapsed). P-uncorrected distances were calculated with MEGA X v10.2.4 [65] on the phylotranscriptomic dataset, whereas patristic distances were obtained from the LG + F + I + I4 + C60 tree.

(g) Relaxed molecular clock

Bayesian molecular dating was performed with MCMCTree [66] within the PAML package v4.9 h [67]. We used the phylotranscriptomic tree (figure 3) and eight fossil calibrations with

uniform, t-cauchy and skew-t prior distributions, following parametrizations in Morris *et al.* [3] (their electronic supplementary material, table S8). We assumed relaxed uncorrelated lognormal molecular clocks (clock = 2) and birth–death tree priors. Analyses used approximate-likelihood calculations [68] on the phylotranscriptomic dataset (single partition) under the LG + I model. A diffuse gamma Dirichlet prior was used for the prior on mean rates as 0.1407 replacements site⁻¹ 10⁸ Myr⁻¹ ('rgene_gamma'; $\alpha = 2$, $\beta = 14.21$). The rate drift parameter reflected considerable rate heterogeneity across lineages ('sigma2_gamma'; $\alpha = 2$, $\beta = 2$). A 100 Ma time unit was assumed. Two independent MCMC chains were run for each analysis, consisting of 22 million generations and the first 2 000 000 were excluded as burnin. Convergence was checked using Tracer v1.7.1 [69]; all parameters reached effective sample size (ESS) > 1000.

(h) Plastid phylogeny

A plastid dataset of 75 proteins [70] was extended by adding the 22 missing species to mimic the phylotranscriptomic nuclear tree (different species of the same genus were sometimes used). Homologous proteins were identified by BLASTP (*e*-value < 1×10^6) from available plastomes or transcriptomes. Genes were aligned with default MAFFT options and trees inferred with IQ-TREE under BIC-selected models and 1000 SH-aLRT. Alignments and gene trees were visualized with FigTree (<https://github.com/rambaut/figtree>) and SeaView [71] to remove paralogues and contaminants. Cleaned gene sequences were masked with PREQUAL, aligned with MAFFT (--allowshift --unalignlevel 0.8), and positions greater than 33% gaps removed with ClipKIT so that final alignments had lengths similar to Ruhfel *et al.* [70]. After concatenation, the plastid dataset consisted of 28 taxa and 16 085 aligned amino acid positions (32% missing data). Maximum-likelihood trees were inferred using IQ-TREE under BIC-selected homogeneous (cpREV + F + I + I4) and mixture (cpREV + F + I + I4 + C60) models and branch support assessed with 1000 UFBoot2 [62] and SH-like aLRT [63] pseudoreplicates.

(i) Quantification of gene expression

Filtered and trimmed reads were mapped against the *Chlorokybus* genome (CCAC 0220) [13] using STAR v2.7.3a [25] (--runMode alignReads --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.05 --alignIntronMin 20 --alignIntronMax 1 000 000 --alignMatesGapMax 1 000 000 --twopassMode Basic --sjdbScore 1 --quantMode TranscriptomeSAM --quantTranscriptomeBan IndelSoftclipSingleend). Gene expression was quantified with RSEM [26] (default parameters), followed by cross-sample normalization (TMM) using edgeR as implemented in Trinity (abundance_estimates_to_matrix.pl). Heatmaps were plotted ComplexHeatmap v2.6.2 in R v4.0.3 [72].

Data accessibility. RNAseq data are available on NCBI (Bioproject PRJNA708203). RNAseq FastQC reports, transcriptome assemblies, preliminary orthogroups, final concatenated alignments, phylogenetic trees, and molecular clock results are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.0gb5mkm25> [73].

Authors' contributions. I.I.: conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing-original draft, writing-review and editing; T.D.: conceptualization, data curation, formal analysis, investigation, methodology, resources, visualization, writing-original draft, writing-review and editing; T.P.: formal analysis, investigation, methodology, resources, visualization, writing-review and editing; J.M.R.F.-J.: formal analysis, investigation; M.J.: formal analysis, writing-review and editing; J.d.V.: conceptualization, funding acquisition, project administration, resources, supervision, visualization, writing-original draft, writing-review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. The authors declare no competing interests.

Funding. This work used the Scientific Compute Cluster at GWDG, the joint data centre of Max Planck Society for the Advancement of Science (MPG) and University of Göttingen. J.d.V. thanks the European Research Council for funding under the European Union's Horizon 2020 programme (Grant Agreement No. 852725; ERC-StG 'TerreStriAL') and the Deutsche Forschungsgemeinschaft (VR132/4-1). J.M.R.F.-J. is supported by the 'Göttingen Graduate Center for

Neurosciences, Biophysics, and Molecular Biosciences' (GGNB), University of Göttingen.

Acknowledgements. We thank Dr Maike Lorenz (SAG, Göttingen) for aid in culture obtainment, Prof. Dr Ute Krämer (RUB, Bochum) for access to the *Arabidopsis halleri* genome and Dr Michael Guiry (NUI, Galway) for help with nomenclatural questions. I.L., J.M.R.F.-J., and J.d.V. are part of the MADLand DFG priority programme 2237 (<http://madland.science>). We thank Debbie Maizels (www.scientific-art.com) for her superb figure 2.

References

- Li L *et al.* 2020 The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat. Ecol. Evol.* **4**, 1–12. (doi:10.1038/s41559-019-1083-z)
- Leebens-Mack JH *et al.* 2019 One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685. (doi:10.1038/s41586-019-1693-2)
- Morris JL *et al.* 2018 The timescale of early land plant evolution. *Proc. Natl Acad. Sci. USA* **115**, E2274–E2283. (doi:10.1073/pnas.1719588115)
- Strasser JFH, Irisarri I, Williams TA, Burki F. 2021 A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879. (doi:10.1038/s41467-021-22044-z)
- Hori K *et al.* 2014 *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* **5**, 3978. (doi:10.1038/ncomms4978)
- Rippin M, Becker B, Holzinger A. 2017 Enhanced desiccation tolerance in mature cultures of the streptophytic green alga *Zygnema circumcarinatum* revealed by transcriptomics. *Plant Cell Physiol.* **58**, 2067–2084. (doi:10.1093/pcp/px136)
- Nishiyama T *et al.* 2018 The *Chara* genome: secondary complexity and implications for plant terrestrialization. *Cell* **174**, 448–464. (doi:10.1016/j.cell.2018.06.033)
- de Vries J, Curtis BA, Gould SB, Archibald JM. 2018 Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc. Natl Acad. Sci. USA* **115**, E3471–E3480. (doi:10.1073/pnas.1719230115)
- Carpenter EJ *et al.* 2019 Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *GigaScience* **8**, giz126. (doi:10.1093/gigascience/giz126)
- Cheng S *et al.* 2019 Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell* **179**, 1057–1067. (doi:10.1016/j.cell.2019.10.019)
- de Vries J *et al.* 2020 Heat stress response in the closest algal relatives of land plants reveals conserved stress signaling circuits. *Plant J.* **103**, 1025–1048. (doi:10.1111/tpj.14782)
- Jiao C *et al.* 2020 The *Penium margaritaceum* genome: Hallmarks of the origins of land plants. *Cell* **181**, 1097–1111. (doi:10.1016/j.cell.2020.04.019)
- Wang S *et al.* 2020 Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* **6**, 95–106. (doi:10.1038/s41477-019-0560-3)
- Wickett NJ *et al.* 2014 Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. USA* **111**, E4859–E4868. (doi:10.1073/pnas.1323926111)
- Irisarri I, Strasser JFH, Burki F. In press. Phylogenomic insights into the origin of primary plastids. *Syst. Biol.* (doi:10.1093/sysbio/syab036)
- Lemieux C, Otis C, Turmel M. 2007 A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biol.* **5**, 2. (doi:10.1186/1741-7007-5-2)
- Fürst-Jansen JMR, de Vries S, de Vries J. 2020 Evolution: On stress responses and the earliest land plants. *J. Exp. Bot.* **71**, 3254–3269. (doi:10.1093/jxb/eraa007)
- Geitler L. 1942 Neue luftlebige Algen aus Wien (Vorläufige Mitteilung). *Österr. Bot. Z.* **91**, 49–51. (doi:10.1007/BF01257345)
- Geitler L. 1942/43 Morphologie, Entwicklungsgeschichte und Systematik neuer bemerkenswerter atmophytischer Algen aus Wien. *Flora NF* **136**, 1–29. (doi:10.1016/S0367-1615(17)31209-0)
- Geitler L. 1955 Über die cytologisch bemerkenswerte Chlorophyceae *Chlorokybus atmophyticus*. *Österr. Bot. Z.* **102**, 20–24. (doi:10.1007/BF01768759)
- Rieth A. 1972 Über *Chlorokybus atmophyticus* Geitler 1942. *Arch. Protistenkd.* **114**, 330–342.
- Seppely CVW *et al.* 2020 Soil protist diversity in the Swiss western Alps is better predicted by topoclimatic than by edaphic variables. *J. Biogeogr.* **47**, 866–878. (doi:10.1111/jbi.13755)
- Dunthorn M *et al.* 2014 Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.* **31**, 993–1009. (doi:10.1093/molbev/msu055)
- Pröschold T, Darienko T, Krienitz L, Coleman AW. 2018 *Chlamydomonas schloesseri* sp. nov. (Chlamydomophyceae, Chlorophyta) revealed by morphology, autolysin cross experiments, and multiple gene analyses. *Phytotaxa* **362**, 21–38. (doi:10.11646/phytotaxa.362.1.2)
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013 STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
- Li B, Dewey C. 2011 RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323. (doi:10.1186/1471-2105-12-323)
- Gontcharov AA, Melkonian M. 2004 Unusual position of the genus *Spirotaenia* (Zygnematophyceae) among streptophytes revealed by SSU rDNA and *rbcl* sequence comparisons. *Phycologia* **43**, 105–113. (doi:10.2216/i0031-8884-43-1-105.1)
- Lax G, Eglit Y, Eme L, Bertrand EM, Roger AJ, Simpson AGB. 2018 Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* **564**, 410–414. (doi:10.1038/s41586-018-0708-8)
- Strasser JFH, Jamy M, Mylnikov AP, Tikhonov DV, Burki F. 2019 New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Mol. Biol. Evol.* **36**, 757–765. (doi:10.1093/molbev/msz012)
- Kawachi M *et al.* 2021 Rappemonads are haptophyte phytoplankton. *Curr. Biol.* **31**, 2395–2403. (doi:10.1016/j.cub.2021.03.012)
- Bremer K. 1985 Summary of green plant phylogeny and classification. *Cladistics* **1**, 369–385. (doi:10.1111/j.1096-0031.1985.tb00434.x)
- Turland N *et al.* 2018 *International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Glashütten, Germany: Koeltz Botanical Books. (doi:10.12705/Code.2018)
- van den Hoek C, Stam WT, Olsen JL. 1988 The emergence of a new chlorophyten system, and Dr. Kormmann's contribution thereto. *Helgoländer Meeresunters.* **42**, 339–383. (doi:10.1007/BF02365617)
- Schlösser UG. 1997 Additions to the Culture Collection of Algae since 1994. *Bot. Acta* **110**, 424–429. (doi:10.1111/j.1438-8677.1997.tb00659.x)
- Schlösser UG. 1994 SAG—Sammlung von Algenkulturen at the University of Göttingen Catalogue of Strains 1994. *Bot. Acta* **107**, 113–186. (doi:10.1111/j.1438-8677.1994.tb00784.x)
- Darienko T, Gustavs L, Pröschold T. 2016 Species concept and nomenclatural changes within the genera *Elliptochloris* and *Pseudochlorella* (Trebouxiophyceae) based on an integrative approach. *J. Phycol.* **52**, 1125–1145. (doi:10.1111/jpy.12481)
- Darienko T, Rad-Menéndez C, Campbell C, Pröschold T. 2019 Are there any true marine *Chlorella* species?

- Molecular phylogenetic assessment and ecology of marine *Chlorella*-like organisms, including a description of *Droopiella* gen. nov. *Syst. Biodiv.* **17**, 811–829. (doi:10.1080/14772000.2019.1690597)
38. Darienko T, Gustavs L, Eggert A, Wolf W, Pröschold T. 2015 Evaluating the species boundaries of green microalgae (*Coccomyxa*, Trebouxiophyceae, Chlorophyta) using integrative taxonomy and DNA barcoding with further implications for the species identification in environmental samples. *PLoS ONE* **10**, e0127838. (doi:10.1371/journal.pone.0127838)
 39. Swofford DL. 2002 *PAUP**. *Phylogenetic analysis using parsimony (*and other methods)*. Sunderland, MA: Sinauer Associates.
 40. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
 41. Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012 A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61**, 973–999. (doi:10.1093/sysbio/sys058)
 42. Gowri-Shankar V, Rattray M. 2007 A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. *Mol. Biol. Evol.* **24**, 1286–1299. (doi:10.1093/molbev/msm046)
 43. Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N. 2013 Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7**, 652–659. (doi:10.1038/ismej.2012.147)
 44. Jamy M *et al.* 2020 Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol. Ecol. Res.* **20**, 429–443. (doi:10.1111/1755-0998.13117)
 45. Mahé F *et al.* 2017 Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **1**, 1–8. (doi:10.1038/s41559-017-0091)
 46. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583. (doi:10.1038/nmeth.3869)
 47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
 48. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013 The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596. (doi:10.1093/nar/gks1219)
 49. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
 50. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
 51. Ju C, Van de Poel B, Cooper ED, Thierer JH, Gibbons TR, Delwiche CF, Chang C. 2015 Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nat. Plants* **1**, 14004. (doi:10.1038/nplants.2014.4)
 52. Haas BJ *et al.* 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512. (doi:10.1038/nprot.2013.084)
 53. Davidson NM, Hawkins ADK, Oshlack A. 2017 SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. *Genome Biol.* **18**, 148. (doi:10.1186/s13059-017-1284-1)
 54. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548. (doi:10.1093/molbev/msx319)
 55. Timme RE, Bachvaroff TR, Delwiche CF. 2012 Broad phylogenomic sampling and the sister lineage of land plants. *PLoS ONE* **7**, e29696. (doi:10.1371/journal.pone.0029696)
 56. Richter DJ, Berney C, Strasser JFH, Burki F, de Vargas C. 2020 EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotic life. *bioRxiv*, 2020.06.30.180687.
 57. Steinegger M, Söding J. 2017 MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028. (doi:10.1038/nbt.3988)
 58. Emms DM, Kelly S. 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
 59. Whelan S, Irisarri I, Burki F. 2018 PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930. (doi:10.1093/bioinformatics/bty448)
 60. Katoh K, Standley DM. 2016 A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933–1942. (doi:10.1093/bioinformatics/btw108)
 61. Steenwyk JL, Iii TJB, Li Y, Shen X-X, Rokas A. 2020 ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biol.* **18**, e3001007. (doi:10.1371/journal.pbio.3001007)
 62. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV. 2017 UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. (doi:10.1093/molbev/msx281)
 63. Anisimova M, Gascuel O. 2006 Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552. (doi:10.1080/10635150600755453)
 64. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018 ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* **19**, 153. (doi:10.1186/s12859-018-2129-y)
 65. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018 MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. (doi:10.1093/molbev/msy096)
 66. dos Reis M, Donoghue PCJ, Yang Z. 2016 Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80. (doi:10.1038/nrg.2015.8)
 67. Yang Z. 2007 PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
 68. dos Reis M, Yang Z. 2011 Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172. (doi:10.1093/molbev/msq317)
 69. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904. (doi:10.1093/sysbio/syy032)
 70. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014 From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **14**, 23. (doi:10.1186/1471-2148-14-23)
 71. Gouy M, Guindon S, Gascuel O. 2010 SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224. (doi:10.1093/molbev/msp259)
 72. Gu Z, Eils R, Schlesner M. 2016 Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849. (doi:10.1093/bioinformatics/btw313)
 73. Irisarri I, Darienko T, Pröschold T, Fürst-Jansen JMR, Jamy M, de Vries J. 2021 Data from: Unexpected cryptic species among streptophyte algae most distant to land plants. Dryad Digital Repository. (10.5061/dryad.0gb5mkm25)

5 Discussion

The study of plant terrestrialization is multifaceted — and many facets remain obscure. One way to shed light on which factors allowed for the success of this fateful event is to look at the conserved stress response mechanisms land plants share with their closest algal relatives, the streptophyte algae. Through comparative studies it is possible to infer the molecular traits that the earliest land plant might have possessed that ultimately aided in the conquest of land. The aim of this thesis was to understand parts of the conserved molecular stress response toolkit that land plants share with streptophyte algae. To investigate into this there have been two top-down approaches, one top-down approach on stress response in two Zygnematophyceae, *Mougeotia* and *Mesotaenium* (Chapter I). The second approach relies on comparative studies of gene sets, building on comparative genomes (Chapter II). The results of both approaches will be discussed in the sections below.

5.1 On Chapter I: A broad view of abiotic stress response in streptophyte algae

One way to explore this toolkit is to screen for stress response in a large-scale setup. Abiotic stress treatments were chosen for all conducted studies. First, these stress environments are easier to setup and to regulate in laboratory conditions than biotic stress applications. Second, when performing a broad and untargeted screen of general algal stress response an objective approach is essential, it is relatively straightforward to apply abiotic stressors in an uniform manner. Finally, deciphering evolutionary conserved traits under biotic stress conditions might be nearly impossible as the interaction between a host and a parasite/pathogen is highly dynamic and characterized by constantly evolving competitive traits.

Various studies have been conducted in land plants as well as some algal systems to probe their response to abiotic stress. Yet, for streptophyte algae, the molecular underpinnings of their stress responses are not well understood. This scarcity in molecular data notwithstanding, streptophyte algae are compelling organisms not only because of their evolutionary connection to the emergence of the first land plants but also because of the remarkable stress resistance of some specimens like *Klebsormidium* or *Zygnema* (e.g., Hartmann *et al.*, 2020; Holzinger *et al.*, 2011; Holzinger *et al.*, 2018; Rippin *et al.*, 2017). The following two chapters of this thesis will shed light on some of these molecular traits that come forth upon a change in environmental conditions and possibly play a role in the conserved stress response mechanism of algae and land plants.

5.1.1 Environmental changes reveal fast-responding adjustment of photosynthetic mechanisms in *Mougeotia* (Publication II)

The earliest land plants likely lived in an everchanging environment. Next to other biotic and abiotic environmental factors, this also included a frequent cycle of wet to dry periods which ultimately also helped in the transition from water to land. But to thrive on land, a flexible and quick response to these environmental changes was crucial for the plant.

Investigating the response to prolonged dry periods (drought/desiccation stress) is the focus of many physiological studies on streptophyte algae (e.g., Holzinger *et al.*, 2011; Karsten *et al.*, 2014; Karsten and Holzinger, 2012). The reverse process, the submergence in water, is however barely investigated. In this study the different species of the Zgynematophyceae *Mougeotia* were submerged in liquid culture medium after growing on solid medium. The focus here was not only on the photo-physiological and morphological response to this change but also on the underlying differential gene expression compared to a control that was not submerged. The results suggest a flexible acclimation of photosynthetic mechanisms in filamentous *Mougeotia* within a short timeframe and will be discussed in the following. For this thesis the center of discussion will be on the observed changes in the *Mougeotia* sp. strain MZCH 240 as most data in Publication II was collected on this strain.

After *Mougeotia* sp. MZCH 240 was grown for 7 days on solid or in liquid medium respectively, the cultures on solid medium were submerged in 10 ml liquid medium for a maximum of 24 hours. Additional Chlorophyll-*a*-fluorescence measurements of dark-acclimated algae were taken at specific time points from which the maximum quantum efficiency of PSII photochemistry (F_v / F_m) was calculated. This served a double purpose. PSII maximum quantum efficiency can be used as a proxy for the physiological status of plant and algal cells (Baker, 2008); concomitantly, this photophysiology in *Mougeotia* was also a focus of this discussed study.

Already after 2 hours the F_v / F_m -values in the submerged cultures were significantly lower than in the control cultures on solid medium, arguing for a fast response to an environmental change in the photosynthetic machinery. A decline in maximal photochemical performance can be interpreted in different ways. If this decrease relates to photoprotective measures like non-photochemical quenching (NPQ) as this process is in direct competition for excitation energy with photochemical quenching remains to be investigated. F_v / F_m values of submerged *Mougeotia* cultures were significantly lower than the control values for up to 8 hours. Then, after 24 hours of submergence, the values adjusted to the control values again, speaking to a recovered photophysiology in submerged *Mougeotia*.

The photo-physiological observations were also reflected in the global differential gene expression profile of submerged *Mougeotia* sp. MZCH 240. Here, RNAseq analyses of *Mougeotia* samples that were submerged for 4 hours as well as control samples only grown on solid medium were carried out and the transcriptome profiles were compared. To gain an overview of the overall gene expression changes, the obtained transcripts were mapped onto the KEGG database (Kyoto Encyclopedia of Genes and Genomes). KEGG pathway maps are particularly apt for capturing photosynthesis-associated alterations in gene expression data. This is due to the high number of the highly conserved set of genes contributing to photosynthesis. In total, 118 KEGG pathways with 347 responsive KEGG orthologs that had a 2-fold change in gene expression level compared to the control could be identified in submerged *Mougeotia* samples (Figure 1 in **Publication II**). The pathways with the highest count in KEGG orthologs included “oxidative phosphorylation [PATH:ko00190]”, “citrate cycle [PATH:ko00020]” and “amino sugar and nucleotide metabolism [PATH:ko00520]”. In these pathways the majority of KEGG orthologs were significantly downregulated, while the pathway “ribosome [PATH:ko03010]” counted the most upregulated KEGG orthologs. The upregulation of the ribosome metabolism, along with the downregulation of amino acid and sugar metabolism, can be interpreted as a general regulatory response as the plant cell has to prepare new/different proteins and therefore needs to produce and transport RNA (see also Pai and Luca, 2019). On the other hand, respiratory processes like oxidative phosphorylation are among the first processes to be downregulated in *Mougeotia*. This speaks of the key role of the mitochondrion in the response to environmental stressors and its retrograde signaling characteristics as one of two organelles that are the main key players for energy balance in the plant cell (Crawford *et al.*, 2018).

On the other hand, the chloroplast is a cornerstone in the overall cell biological status of plant cells (Gläßer *et al.*, 2014) — and thus likely of algal cells, too. This is reflected in the fast upregulation of photosynthesis-related pathways in *Mougeotia*. Here, pathways relating to photosynthetic processes (“Porphyrin and chlorophyll metabolism [PATH:ko00860]”; “Photosynthesis [PATH:ko00195]”) were among the pathways with the most differentially expressed KEGG orthologs. In the “Porphyrin and chlorophyll metabolism” pathway, mainly orthologs from genes that are involved in the biosynthesis of protoporphyrin IX, a precursor for chlorophyll biosynthesis (gene names in KEGG: hemE, hemC, hemA, hemL) (see also Sachar *et al.*, 2016) are upregulated, speaking to a possible adjustment of chlorophyll content in *Mougeotia*. Interestingly, such chlorophyll precursors have been suggested to be integral to retrograde signaling (Koussevitzky *et al.*, 2007); however, this notion is not without controversies (Mochizuki *et al.*, 2008).

The “Photosynthesis” KEGG pathway mainly showed a differential regulation of PSI and PSII genes (17 photosystem I and photosystem II subunits out of 23 KEGG identifiers for this pathway, see Figure 1, **Publication II**). This readjustment of core genes of the photosynthetic apparatus as well as chlorophyll biosynthesis, is reasonable considering *Mougeotia* is confronted with a sudden change of light availability and light quality due to submergence. This adjustment could therefore also explain the significant drop of the photosynthetic yield that was already observed after 2 hours of submergence.

To address specific gene expression changes upon submergence, a BLAST-based homology (BLAST: Basic Local Alignment Search Tool) search against the *Arabidopsis thaliana* genome obtained from TAIR was performed. This search revealed 120 significantly (Benjamini-Hochberg corrected $p \leq 0.001$, differential gene expression changes of at least 2-fold) and 171 significantly downregulated genes. Upregulated genes reflected the results previously obtained through analysis using KEGG orthologs and revealed a response in photosynthesis-related processes upon submergence. Several genes responsible for chlorophyll a/b-binding proteins and were upregulated (see Table 2 or Figure S2, **Publication II**). Furthermore, the upregulation of genes responsible for pigment synthesis, especially carotenoids, was observed. These include a homolog of *ABA3*, coding for the cytosolic molybdenum cofactor sulfurase, involved in the biosynthesis of the phytohormone ABA by converting carotenoid-derived abscisic aldehyde in land plants (Bittner *et al.*, 2001). Whether *Mougeotia* *ABA3* is involved in ABA biosynthesis remains obscure, yet it is likely that this gene acts in the framework of the highly conserved carotenoid biosynthesis pathway. Additionally, a violaxanthin de-epoxidase and a carotenoid cleavage dioxygenase (a homolog of *CCD1*) were upregulated. These changes in expression of genes for pigment metabolism underline again that a photosynthetic organism needs to quickly readjust structures involved in light harvesting when light quality and quantity suddenly change to uphold vital photosynthetic performance.

After 24 hours of submergence, the morphology of *Mougeotia* sp. MZCH240 was assessed through light micrographs (Figure 3, **Publication II**). Here, the liquid-grown control samples showed the most notable phenotypes compared to the submerged culture and the solid control. Rhizoid formation could be observed here, which had been observed before for *Mougeotia scalaris* (Buschmann, 2020) and has been described multiple times for the Zygnematophyceae *Spirogyra* (Yoshida *et al.*, 2003; Yoshida and Shimmen, 2009). Additionally, the liquid-grown *Mougeotia* filaments showed inclusions with brownish pigment. Whether this pigment is a phenolic compound is unclear, although phenolic compounds have been reported in other Zygnematophyceae multiple times (Holzinger *et al.*, 2018; Aigner *et al.*, 2013; Pichtrová *et al.*, 2013).

Additionally, phenolic compound was detected in *Mougeotia* before where they originate from vesicles near the chloroplast and were hypothesized to play a role in light-oriented chloroplast movements (Schönbohm and Schönbohm, 1984). The chloroplast movement in *Mougeotia* is a phytochrome-mediated photo-modulated process that is especially observed under high irradiance (Grolig and Wagner, 1988) and was also observed in the discussed study (Figure 3, **Publication II**). The formation of these brown inclusions, together with the reorientation of chloroplasts, could possibly be part of a response mechanism to the environmental conditions, but this must be investigated further to give a clear statement.

Lastly, the formation of storage granules could be observed in the control cultures as well as the submerged *Mougeotia* filaments. Whether these structures are land plant like lipid droplets is, for now, unknown. But the obtained global gene expression profile of *Mougeotia* can give a first insight into this question. Indeed, transcripts coding for land plant hallmark lipid droplet associated proteins like PUX10, LDAP3 or HSD1 could be recovered from the transcriptome data set of *Mougeotia sp.* MZCH240 (suppl. Table 2, attached supplement at the end of this thesis). Here the BLAST search was not only conducted against known data from the land plant *Arabidopsis thaliana*, but from the Zygnematophyceae *Mesotaenium endlicherianum* strain SAG 12.97 as well, resulting in a recovery of more lipid droplet associated gene IDs than in the BLAST search against *A.thaliana*. As a lot of global gene expression data, including a proteomic profile, was created for *Mesotaenium* SAG 12.97, detecting also orthologs for lipid droplet associated proteins (more on this in the next chapter, **Publication III**), this dataset was chosen as an additional reference here. To scrutinize if the structures observed in *Mougeotia sp.* MZCH240 are indeed land plant like lipid droplets, an analysis on TAG content or/and a proteomic analysis must be conducted in the future.

5.1.2 In-depth multifaceted analysis reveals plastid-derived molecular programs in *Mesotaenium* (Publication III)

Terrestrialization went hand in hand with a combination of different stressors. To survive, plants must possess an elaborate and fast-responding molecular network. In land plants underlying genetic networks that act under a combination of (i) different abiotic stressors as well as (ii) abiotic co-occurring with biotic stressors are well studied (see e.g., Desaint *et al.*, 2021). To understand how this conserved stress responsive toolkit might have facilitated plant terrestrialization, it is important to not only look at stress response in land plants. One needs to turn to the genetic capacities of land plants' closest algal relatives (the Zygnematophyceae) and infer the toolkit that is shared between them; this sheds light on the molecular chassis that was present in the last common ancestor of land plants and zygmatophyceae algae.

In recent years there have been advances in understanding the stress-responsive networks in streptophyte algae (e.g., Jiao *et al.*, 2020). Nevertheless, a lot still needs to be uncovered. Therefore, we conducted a large-scale study with the aim of giving insight into the molecular mechanisms that play a role under an assault of combinatory abiotic stressors in one representative of the Zygnematophyceae.

In Dadras *et al.* (2023; **Publication III**), a unique laboratory setup (a ‘gradient table’) was used to create a bifactorial gradient of temperature and irradiance, two major terrestrial stressors. These stressors were applied to the Zygnematophyceae *Mesotaenium endlicherianum* (strain SAG 12.97). This freshwater/subaerial dwelling alga was chosen for multiple reasons, the foremost reason being the availability of its genome (Cheng *et al.*, 2019). Using the genome data as a backbone facilitated robust analysis of the obtained transcriptome data in this study which was then mapped on the published genome. Furthermore, *Mesotaenium* is mostly unicellular and therefore easy to keep in a homogenous culture, which is favorable when performing (photo-) physiological analyses as it increases their accuracy by allowing for an equal distribution of environmental cues such as light to all cells. Another advantage is that the strain SAG 12.97 is axenic, meaning there are no bacterial and/or fungal contaminations. Lastly, for this large-scale setup a generous amount of culture was needed, which is easy to implement as *Mesotaenium* grows fast to an adequate cell density—it took, on average, 12 days in liquid culture with a carbon source. Starting with the right number of cells is crucial for stress analysis, especially involving high irradiance, as a high number of cells can lead to self-shading, thereby protecting lower layers of cells from the irradiance and introducing an unwanted factor in stress response analysis. On the other hand, a low number of cells can equally impede further analyses as, depending on the intensity and the duration of stress, too few cells will be available for downstream analyses. Thus, pre-studies were performed with the aim of fine-tuning the experimental setup.

Every organism responds to stress differently. Therefore, for this study the conditions chosen were entirely adjusted to induce a stress response in *Mesotaenium* SAG 12.97 that was still analyzable. This adjustment included setting the temperature gradient from 11.6 °C to 34.7 °C to a new gradient ranging from 8.6 °C to 29.2 °C. Here two parts of the analysis were considered a bottleneck and therefore taken as a reference to adjust culturing as well as stress conditions. For one, this was the photo-physiological assessment of stress-treated cultures using pulse-amplitude modulation (PAM). Here chlorophyll-*a*-fluorescence is only measurable if a certain number of cells is still vital, or at least still contain detectable chlorophyll *a*. The other factor is the isolation of RNA for down-stream analysis of the transcriptome and is also strongly determined by the cell density and cell vitality at the end of the stress treatment.

The experimental setup tailored to the *Mesotaenium* strain SAG 12.97 resulted in a fine grid of a gradient consisting of two terrestrial stressors, irradiance (ranging from 21.0 to 527.9 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) and temperature (ranging from 8.6 °C to 29.2 °C), on a 2D gradient table (see Figure 1b, Extended Data Figure 4a, b, c in **Publication III**). 42 twelve-well plates containing liquid *Mesotaenium* cultures were put on this table for 65 hours and were afterwards analyzed regarding their morphology, (photo-) physiology and global gene expression profile, which was the prime focus of this study. Three successive biological replicates of this setup resulted in the multi-faceted analysis of 126 samples, the results of which will be discussed below.

The physiological phenotype the *Mesotaenium* cultures displayed was strongly dependent on the temperature and irradiance conditions. With rising light intensities F_v/F_m as well as absorption values (at 480, 680 and 750 nm) decreased with F_v/F_m -values reaching down to zero at the highest irradiance on the gradient (527.9 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) coupled with either the lowest temperature (8.6 °C) or the highest temperature (29.2 °C). Low temperature coupled with rising irradiance had a significantly ($p \leq 0.001$) stronger negative impact on growth and photo-physiology than high temperature. Here F_v/F_m -values measured at low temperature form a large significant group that differs ($p \leq 0.001$) from the significant group under the same light but high-temperature conditions (see Figure 1d, **Publication III**). The additional clustering of the physiological data against either measured irradiance or temperature values reveals a less broad distribution if clustered by irradiance than by temperature (Figure 1e, 1f, **Publication III**). This, along with the observed patterns independent of clustered values (Figure 1c, **Publication III**), could imply a broader tolerance for temperature (eurythermy) which in turn then facilitates a tolerance for high irradiance conditions (euryphoty).

For global gene expression profiling, RNA was isolated out of 114 *Mesotaenium* samples, representing 38 conditions. After data processing, a broad overview of the general gene expression profile of the samples was obtained by using principal component analysis (PCA) as well as Euclidean distance analysis and Spearman correlation (Figure 2a, 2b, 2c, **Publication III**). While the PCA revealed that independent biological replicates cluster together in proximity, the distance analyses could separate the clusters in three main categories: 1) high irradiance and/or high-temperature samples, 2) low-temperature samples from 8-17°C and 3) a large “steady state” sample cluster. The term “steady state” is here defined as the environment in which optimal growth conditions and low observable stress for *Mesotaenium* dominate, which we assessed based on high absorption as well as photosynthetic performance values of the alga. A cluster representing high irradiance coupled with low temperature samples is missing here (and in the data sets discussed below) because RNA extraction was not successful for samples exposed to these conditions.

To get an overview over differential gene expression patterns between the samples they were divided into nine sectors plus one cohort with samples that showed F_v/F_m -values below 0.5, which was defined as a stressed set of algal samples, and were compared against each other. Differentially expressed genes of the 36 performed comparisons (see Figure 2g, **Publication III**) were only considered at a fold change ≥ 2 and were Benjamini-Hochberg corrected ($p \leq 0.001$). Overall comparisons with the highest number of differentially expressed genes were also characterized by opposite environmental conditions like low light coupled with low temperature (LLI_LT) versus high light and high temperature (HLI_HT), which is reasonable since it is conceivable that a genetic response adapted to the prevailing condition is to be expected for the alga to survive. To understand which genes were significantly upregulated collectively in all/most of the sectors where the samples were considered “stressed”, an enriched biological theme comparison coupled with a GO-term enrichment analysis was performed, using the low light, middle temperature sector, the “steady state” condition, as a control to compare the other conditions against, which resulted in 9 comparisons (Extended Data Figure 1, **Publication III**). These analyses revealed the enrichment of plastid-biology related genes in all sectors compared to the “steady state” condition. Furthermore 30 genes that were upregulated in all 9 comparisons revealed GO-terms for multiple genes coding for chlorophyll A/B binding proteins and photosystem II light harvesting complex proteins (Extended Data Figure 7b, **Publication III**). These results are not surprising as plastids are major players in the initial cellular response to environmental changes in land plants and recapitulate observations made for other stressed streptophyte algae, including *Mougeotia*.

The screen of differential gene expression between the nine environmental sectors revealed a shared response among analyzed sectors speaking of underlying genetic programs that operate under a broad range of environmental conditions in *Mesotaenium* SAG 12.97. To investigate these genetical programs a weighted gene co-expression network analysis (WGCNA), an unsupervised clustering method that clusters genes based on their expression profiles, was chosen. The strength of the unsupervised clustering in this context was that the analysis is objective and independent from homology verifications to land plant genes, informing us about the molecular programs that operate in the alga without bias or preconceptions. The result of the WGCNA were 26 clustered modules, for which different colors were assigned for orientation, and which encompass 17,905 analyzed genes in total (Figure 4, **Publication III**). These modules were then correlated with physiological as well as environmental conditions (Figure 4a, b, c, **Publication III**). The first observation is that physiological values tend to correlate negatively with rising light intensities, while this negative correlation was not as strongly pronounced when correlating these values with varying temperatures (Figure 4b, **Publication III**).

This discovery reflects the finding of the photo-physiological analysis and undermines the theory, that this *Mesotaenium* strain might be indeed eurythermic. How the clustered modules might play into that and what other underlying genetic programs they can reveal is discussed in the following paragraphs.

An established and efficient plastid-to-nucleus communication is essential for survival under fast changing environmental conditions in land plants (Cheng *et al.*, 2016; Zhao *et al.*, 2019). Even though there is an ongoing debate about the evolutionary conservation of retrograde signalling cascades in streptophytes (Honkanen and Small, 2022), patterns of retrograde signalling and a variety of plastid-derived signals were found in the light of this study. The blue cluster module (see Figure 4, **Publication III**), correlates negatively with increasing light and positive with high photo-physiological values (F_v/F_m , absorption at 480 and 680 nm) and belongs to the modules with the highest genes counts. A GO-term enrichment analysis of this module showed a high number of GO-terms for cellular signalling, plastid-related processes combined with signalling processes, speaking to signalling coming from the chloroplast (Extended Data Figure 3d, **Publication III**). Among the most connected hubs in this cluster were genes for transcriptional factors (TFs) involved in retrograde signalling in land plants (see Figure 5a, **Publication III**). These included a homolog of GLK1 (Golden-like 1) which is the second most connected hub in blue, and most connected TF. GLK1 plays a role in chloroplast biogenesis, supporting chloroplast-nucleus communication in the expression of photosynthetic machinery related compounds (Waters *et al.*, 2009; Rossini *et al.*, 2001). Damaged chloroplasts inhibit GLK1 expression in *Arabidopsis* (Martin *et al.*, 2016), therefore the negative correlation with rising light intensities of the blue module in this study seems reasonable. Furthermore, a significant change in expression of GLK1 under high temperatures was observed in *Mougeotia* before (de Vries *et al.*, 2020) and therefore seems to be present not only in *Mesotaenium* but in other Zygnematophyceae like *Zygnema* (see **preprint VII**), *Mougeotia* and *Spirogyra* as well, speaking of a possible conserved mechanism in Streptophyta that deserves further attention in the future. Furthermore, a homolog of COL3 (CONSTANS-like 3), which in land plants is also involved in plastid-nucleus-communication was found in the blue cluster as well (4th most connected TF). COL3 is a positive regulator of red light signalling during early photomorphogenesis (Liu *et al.*, 2021). The *MeCOL* homologs (*MeCOL* a,b,c,d) fall into a phragmoplastophyte clade with the *A.thaliana* COLs (see Figure 5a, **Publication III**). Another homolog that was present in the blue module is COP1 (CONSTITUTIVE PHOTOMORPHOGENIC 1). COP1 codes for a E3 ubiquitin ligase which degrades both COL3 and GLK1 and is a negative regulator for photomorphogenesis in *Arabidopsis* (Osterlund *et al.*, 2000; Saijo *et al.*, 2003). Therefore, it is reasonable that this hub shows up in the blue cluster, which correlates negatively with rising light intensities.

Other clustered modules that fall under the umbrella of plastid-derived signals are e.g., the green module, the yellow module, or the light cyan module. While the green and light cyan modules contain GO-terms related to stress response under ROS influence, the yellow cluster contains GO-terms related to protease responses, which are possibly also plastid-related (Extended Data Figure 3a, **Publication III**). All three modules correlate positive with rising light intensities while at the same time showing a negative correlation to photo-physiological values. Notable highly connected hubs in the yellow cluster are involved in chloroplast development and chloroplast protein homeostasis like CLPPs/CLPRs (caseinolytic proteases / caseinolytic proteases-related) (among the most connected hubs are CLPP3, 4, 6), GUN2 (GENOMES UNCOUPLED2) and pTAC6 (PLASTID TRANSCRIPTIONALLY ACTIVE 6) (see Figure 5b, **Publication III**). CLPP/CLPR serine proteases are highly conserved proteases that are involved in plastid protein maintenance in plants (Adam *et al.*, 2006). In *Arabidopsis* no short-term expression level change could be detected for chloroplast Clp proteins, however, long-term high light and cold exposure led to an increase in mRNA and protein content for ClpD and ClpP (Zheng *et al.*, 2002). A phylogenetic analysis revealed that individual MeCLPPs and MeCLPRs fall into one clade with *A. thaliana* CLPP and CLPRs. GUN genes are major players in retrograde signaling with GUN2 being directly involved in the tetrapyrrole biosynthesis pathway (Wu and Bock, 2021). An homolog of GUN1, which participates in multiple plastid signaling pathways (see e.g. Shimizu and Masuda, 2021), was also detected. *MeGUN1* falls into one phragmoplastophyte clade with *A. thaliana* GUN1. pTAC6 is responsible for plastid gene expression in *A. thaliana* and mediates phytochrome signaling (Liebers *et al.*, 2020). Worth mentioning is also the presence of a *Mesotaenium* homolog of HY5 (ELONGATED HYPOCOTYL 5) in the yellow cluster, a transcription factor that is a master regulator of light signalling in land plants (Ang *et al.*, 1998) and gets activated under rising light intensities reasoning well with the positive correlation of the yellow cluster with rising light intensities.

The green module additionally revealed a strongly connected hub was CGI-58 which also fell into one clade including Streptophyte CGI-58 (see Figure 5c, **Publication III**). This enzyme plays an important role in lipid homeostasis in land plants and Lipid droplet accumulation was observed in *Arabidopsis* plants deficient in this enzyme (James *et al.*, 2010). This is especially intriguing considering that lipid droplet formation was also observed in this study in *Mesotaenium* under conditions that correlate also with low photo-physiological values (varying temperatures in combination with high irradiance). This important finding will be discussed at a later point.

Besides modules strongly relating to plastid-derived signalling, two more modules are worth mentioning here: the purple and the pink module. The purple module comprises enriched GO-terms that describe cell division processes. This module correlates with low irradiance intensities and simultaneously high photo-physiological values. *Mesotaenium* samples in these conditions therefore

seem to perform well, and underlying molecular programs seem to facilitate growth. An interesting hub that was found through the network analysis was a kinesin homolog of proteins such as POK2 (PHRAGMOPLAST ORIENTING KINESIN 2) (Figure 5e, **Publication III**) which is involved in correct cell wall positioning in *Arabidopsis* (Chugh *et al.*, 2018) and is likely conserved across Phragmoplastophyta (see Figure 5e, **Publication III**). The gene expressed in the pink module on the other hand, are involved in cell wall-derived signals as the analysed GO-terms suggest. This module correlates as the purple model negative with light and positive with physiological values but the difference to the purple cluster is that gene expression of this module is more pronounced in high temperatures.

Lastly, this study uncovered also an intriguing morphological observation in *Mesotaenium* SAG 12.97: an accumulation of structures that resemble land plant-like lipid droplets (LDs) (Figure 6, **Publication III**). The formation of lipid droplets is common as an answer to a variety of stressors in land plants, with the formation of LDs in seed being among the most researched topics in that field because of economic relevance (e.g., Gasulla *et al.*, 2013). On the other hand, there are, as of today, no studies that investigated land plant-like lipid droplet formation in streptophyte algal classes in greater molecular detail, even though the underlying genetic repertoire for those structures seems to be conserved among streptophytes (de Vries and Ischebeck, 2020). Often, these structures are morphologically observed in streptophyte algae and dubbed as storage granules or lipid bodies but are neither stained, nor isolated to investigate components (**Publication II**; de Vries *et al.*, 2020; Pichrtová *et al.*, 2016). As part of this study, the identity of these structures in long-term stressed, possible nitrogen-starved *Mesotaenium* SAG 12.97 cells, was confirmed via confocal microscopy as well as proteomic analysis (Figure 6a, b, d, **Publication III**).

For confocal micrographs lipid droplet-like structures in *Mesotaenium* were successfully stained with the neutral lipid stain BODIPY™ 493/503 (EM/EX). Additionally, a lipid enriched fraction was isolated from the same stressed cultures the confocal analysis was performed on. The proteomic analysis of these isolates compared to a total extract revealed homologs of hallmark lipid droplet proteins known in land plants that were significantly enriched in the lipid fraction of the isolates, like the cycloartenol synthase (CAS), three putative hydroxysteroid dehydrogenases (HSD) homologs, two oleosin (OLE7) homologs and a homolog for a LD-associated protein (LDAP3). For LDAP3 a phylogenetic analysis revealed that *Mesotaenium* LDAP3 homolog (new gene model: Me1v20056100.1, gene model from Cheng *et al.*, 2019: MesotaeniumME000109S10929) falls into one clade together with other streptophyte algae candidate genes like *Klesormidium nitens*, *Spirogyra pratensis*, *Zygnema circumcarinatum* (more on the genetic landscape of *Zygnema* in **preprint VII**) or *Chara braunii* (Supplementary Figure 28, **Publication III**), an observation that deserves further investigation in future studies. Another interesting observation is the volcano plot analysis of the

proteomic data that revealed significantly enriched *Mesotaenium* protein identifiers (Me1_v2_0114060.1, Me1_v2_0204200.1 and Me1_v2_0185210.1, see also Figure 6d, **Publication III**) for which no *Arabidopsis* BLASTp hit could be found. Interestingly for one *Mesotaenium* identifier a BLASTp hit for a *Klebsormidium nitens* identifier that codes for a putative S-adenosyl-L-methionine-dependent methyltransferase could be identified (suppl. Table 4, for an overview, see also suppl. Table 3, both attached supplement at the end of this thesis). This type of enzyme can possibly play a role in lipid droplet formation (Borrego *et al.*, 2021), but further analyses are needed to make a clear statement here. To further strengthen these proteomic analyses the triacylglycerol (TAG) content should be analyzed in future studies, as TAG content is another important characteristic for land plant-like lipid droplets.

Tying back to the environmental gradients applied to *Mesotaenium* in this study, transcripts for lipid droplet-associated genes showed differential expression profiles under the various environmental conditions (Figure 6c, **Publication III**). This reveals a fine-tuned regulation of these components as their expression profiles are unique for all 38 analysed environmental conditions. While genes like HSD1 or OLE7 mainly showed an upregulated expression under higher temperatures, LIME1 and LDPS (lipid droplet methyltransferase, lipid droplet protein of seeds) were upregulated in both high and low-temperature conditions. The expression of some of these genes was further enhanced under high temperatures in combination with high irradiance (for PUX10 homologs, HSD1, OLE7, LIME2). If that same phenomenon applies to low temperatures as well remains to be investigated, as the gene expression profile of samples exposed to low temperature combined with high irradiance could not be analyzed. Some genes also showed differential expression profiles under “steady state” conditions [CAS, LDIP (LDAP interacting protein), LDAH1 (lipid droplet-associated hydrolase)], hinting at a general regulatory program independent on stress-inducing conditions in *Mesotaenium*.

All in all, these intriguing latest findings in *Mesotanium* open up several routes for future investigations as new questions keep emerging. Lipid droplet formation in streptophyte algae remains mysterious and an interesting topic especially considering their evolutionary conservation and possible role during the conquest of land as a general stress response mechanism.

5.1.3 Concluding remarks on Chapter I

The two studies discussed above show that plastid-derived signals are at the heart of abiotic algal responses to abiotic stress. As photosynthesis plays a vital part in every plastid-bearing organism, this finding does not come as a surprise. A change in environmental conditions often has a direct and rapid influence on photosynthetic performance. Since the plant cell is highly dependent on the photosynthetic apparatus a fast response is only reasonable. The first study showed exactly that: a fast adjustment of photo-physiological components in the filamentous *Mougeotia* as response to a change in light quantity and quality. This included changes in differential gene expression of light harvesting properties, first and foremost chlorophyll biogenesis and carotenoid metabolism. Furthermore, *Mougeotia* showed promising morphological phenotypes upon an environmental change which deserve further investigation in the future.

The second study confirmed the importance of plastid-derived signals as an answer to changing conditions and added onto the findings of the first study by providing deeper insights into underlying co-expression networks of molecular programs in the unicellular alga *Mesotaenium* which also hinted at evolutionary conserved plastid- and cell wall-derived signaling cascades being part of the adjustment to a change in environmental conditions in Zygnematophyceae. The presence of land-plant like lipid droplets in *Mesotaenium* along with the underlying genetic repertoire coding for hallmark land plant LD-associated proteins hints at a possible role of this conserved stress response mechanism in the conquest of land.

5.2 On Chapter II: Comparative genomics can uncover conserved genetical features facilitating terrestrialization

Comparative genomics are crucial to understand molecular plant responses that aided in plant terrestrialization. They can give insights on conserved molecular mechanisms that are present across the green lineage, shedding light on underlying crucial signaling cascades and the biosynthesis of compounds that are relevant for plant survival (see **Publication I**, **Publication IV**, and **Publication VI**). While individual stress screens on organisms of interest can give novel insights, only comparative genomics can set these findings in an evolutionary context.

A comparative, often purely bioinformatic, approach can rely on a transcriptomic, stress responsive analysis. Unique gene expression profiles induced under diverse conditions help to further enhance the genetic landscape of the organism of interest, especially if there is a limited amount of genetic information accessible. For non-model plant organisms like the ones analyzed in this thesis, this approach is a powerful method to garner a breadth of insights and create the foundation for comparative analyses-based inferences.

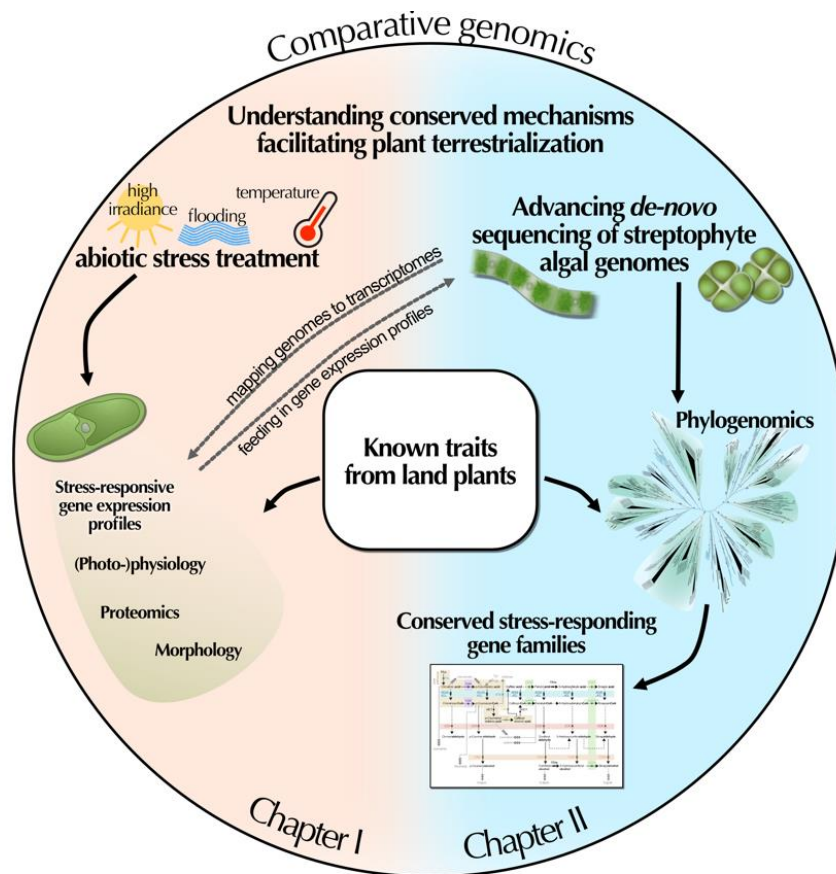


Figure 8: Unraveling conserved mechanisms with comparative genomics. This simplified scheme shows how the two approaches of this thesis (Chapter I, Chapter II) are intertwined and can improve data analysis when used together. The overall aim is to understand conserved mechanisms that play a role in terrestrialization. The point of reference for both approaches are the data sets revealing conserved traits that are known in land plants. Only a comparison with these data sets can infer the conserved traits that streptophyte algae and land plants share and that were therefore present in the earliest land plants (see also Figure 6).

On the other hand, untargeted stress response analyses like they were performed in the light of this thesis can profit greatly from previously performed comparative genomic analyses as well (Figure 8).

How untargeted stress response analyses in streptophyte algae profit from a comparative approach was shown in **Publication III**. The presence of a published *Mesotaenium* SAG12.97 genome (Cheng *et al.*, 2019) was one of the reasons why this strain was chosen for this large-scale setup. Mapping to an already present genome provided a significant boost in quality of per-gene inferences and global stress induced transcriptomic profile. At the same time, it could be illustrated how these data from transcriptome dynamics feed back onto the static genome; based on the RNAseq data we have updated the gene models. The advance of *de-novo* sequencing of other streptophyte algae genomes in recent years has been a huge help in understanding streptophyte algae genetics and pinpointing mechanisms, that are shared across the individual classes and across the whole clade of Streptophyta.

The three studies discussed below will now give an insight on how a comparative approach can be applied in different biological contexts. These studies all have a different focus, from investigating a specific set of stress responsive gene families, to uncovering the evolution of signaling networks, to deciphering streptophyte algae diversity, but the overall aim of all these studies is to understand streptophyte evolution and further unravel the mystery of plant terrestrialization. The next discussed study will be of special interest for the latter. Here comparative phylogenomics decipher the mysteries of the evolutionary conservation of a large set of gene families, that play a major role in land plant stress response and are hypothesized to be major players in plant terrestrialization.

5.2.1 The deep evolutionary roots of the phenylpropanoid pathway (Publication V)

Specialized metabolites define the biology of land plants by influencing growth, development, and physiology, especially regarding a response to changing environments. In light of plant terrestrialization it is therefore even more fascinating that streptophyte algae already possess homologs of enzymes that constitute the routes of these special metabolic pathways (**Publication IV**). One of these important land plant-characteristic pathways is the phenylpropanoid pathway.

The tight involvement of this specialized pathway in diverse stress response mechanisms across the streptophyte lineage makes it an intriguing candidate for studies on the underlying molecular chassis that enabled the earliest land plants to conquer land. It might have provided the earliest land plant with an acclimated set of stress-relevant compounds (Rensing, 2018). The reason for this major role in plant terrestrialization lies in the underlying enzymatic capacity, the driving force behind the highly specialized array of compounds that ward off environmental stressors.

Specialized metabolism is shaped by the environment and is more prone to changes than the evolutionary constricted primary metabolism (**Publication IV**). Aside from environment-specific products, the underlying enzymatic network producing those compounds can be adapted, allowing for high flexibility and an enormous variation of, possible stress-relevant, product formation under specific conditions. Enzymes of the phenylpropanoid-core pathway show this lineage-specific promiscuity which is shaped by the environment. Enzymatic promiscuity describes the ability to catalyze side reactions besides the main reaction. This catalytic reaction happens under continuous selective pressure, for instance, environmental changes, and can provide a fitness advantage (see also Khersonsky and Tawfik, 2010; Copley, 2017). Promiscuity can hence tell us something about how complex pathways with myriads of reactions can evolve over time.

The study discussed below highlights the evolution of some of the major enzyme families that characterize the phenylpropanoid pathway by applying phylogenetic and sequence analyses. This thesis will hereby focus mainly on the enzyme that catalyzes the first step leading into the phenylpropanoid pathway, PAL (Phenylalanine ammonia lyase), and discuss examples of the evolution of other important enzyme families included in phenylpropanoid synthesis briefly.

The L-phenylalanine ammonia lyase (PAL) converts phenylalanine into cinnamate. Next to PAL there is also the bifunctional PTAL (L-tyrosine ammonia-lyase) which converts L-tyrosine into *p*-coumarate and which is present in monocots (Barros and Dixon, 2020). Genes coding for this enzyme were long thought to be land plant-specific (Emiliani *et al.*, 2009), but the recent advances in streptophyte algae *de-novo* genome sequencing revealed promising candidates for this enzyme, first and foremost in the alga *Klebsormidium nitens* (de Vries *et al.*, 2017). The screen performed in this study, which comprises taking Arabidopsis PAL (*AtPAL1*) as bait and screening via BLAST analysis for homologous genes across the green lineage, confirmed the *K. nitens* genes as a promising PAL candidate (Gene Id: kfl00104_0290_v1.1, 479 aa (amino acids) long) (Figure 2, **Publication V**). Next to the *Klebsormidium* PAL candidate, two PAL homologs in *Chara braunii* were also found (g57646_t1 and g34530_t1, 527 aa). Maximum likelihood phylogenetic analyses however revealed that these *C. braunii* PAL homologs grouped with diverse prokaryotic and eukaryotic HALs.

Predictions of protein structures via I-TASSER were performed for the putative *K. nitens* PAL (kfl00104_0290_v1.1) as well as the putative *C. braunii* HAL (g34530_t1); the best structural hits were *Petroselinum crispum* PAL and *Pseudomonas putida* HAL, respectively. For the putative *Chara* HAL even though the best structural hit was *Pseudomonas putida* HAL, the predicted ligand was tyrosine (C-score 0.80). The reported active side chain residues from structural analyses of *Sorghum bicolor* PAL (*SbPAL*, Jun *et al.*, 2018) and *Petroselinum crispum* PAL (*PcPAL*, Nagy *et al.*, 2019) were compared with the predicted residues from the I-TASSER analysis as well as from the performed alignment of all analyzed sequences (Data S1, **Publication V**). Here residues from the short putative PAL sequence in *Mesotaenium* matched with the *SbPAL* and *PcPAL* in 6 residues (Y9, L35, S38, H39, K150 and I154). The putative *Klebsormidium* PAL matched in all residues that are involved in forming the binding pocket, except for Phe, which is substituted by Gly (position 114, Data S1, Figure 2, **Publication V**). This exchange can determine PAL and PTAL function (Jun *et al.*, 2018). In the putative *Chara braunii* PALs Phe is still present. PAL activity hinges on the prosthetic electrophilic group 4-methylideneimidazole-5-one is formed post-translationally (MIO; Schwede *et al.*, 1999). It is a critical and characteristic catalytic group in the binding pocket of PAL. This amino acid triad that becomes the MIO group was conserved in both *Chara* and *Klebsormidium*.

The phylogenomic analysis confirmed the *K.nitens* sequence kfl00104_0290_v1.1 as a putative PAL sequence as it was in accordance with earlier analysis (de Vries *et al.*, 2017) among bacterial PALs (Figure 2, **Publication V**). All in all, there must be further analysis to scrutinize the origin of streptophyte PAL as the phylogenetic analysis reveals a split between embryophyte PALs and the putative *Klebsormidium* PAL, which is nestled in the bacterial PAL clade. Explanations for this scattered evolutionary pattern can be gene losses or specific domain acquisitions, or a horizontal gene transfer that happened in the ancestor of land plants during symbioses with soil bacteria (Emiliani *et al.*, 2009).

Next, the evolution of the cinnamate 4-hydroxylase (C4H), which converts cinnamate to p-coumarate, was investigated through a phylogenomic analysis. C4H belongs to a subfamily of of CYP450 enzymes (CYP73). No clear homolog coding for the monooxygenases, to which also C4H belongs, could be found in the streptophyte algae genome/transcriptome data available. A C4H-independent route to the synthesis of p-coumarate is therefore suggested since p-coumarate could already be detected in core chlorophytes (Goiris *et al.*, 2014).

Besides the conversion to p-coumarate, cinnamate can also be converted into cinnamoyl-CoA by the 4-Coumarate-CoA ligase (4CL) and potentially other enzymes annotated as acyl-CoA synthases (ACS/ACoS, Shockey *et al.*, 2003). Here the phylogenetic analysis revealed that the common ancestor of land plants likely possessed one ACOS5-like and one 4CL-like gene, as both families expanded later during land plant evolution in for instance *Physcomitrium* or *Arabidopsis* (Figure 3, **Publication V**). Several potential streptophyte algal 4CL homologs could be recovered and confirmed by I-TASSER for *Klebsormidium nitens* and the Zygnematophyceae *Spirogyra* and *Zygnema* indeed suggesting the presence of a 4CL/ACOS5-like encoding gene in the last common ancestor (LCA) of streptophytes.

A last excerpt of the discussed study is the evolutionary trajectory of two enzymes involved in lignin-biosynthesis. The biogenesis of the structural polymer lignin involves two successive steps catalyzed by CCR (Cinnamoyl-CoA reductase) and CAD (cinnamoyl alcohol dehydrogenase). The evolution of CCR/CCRL (Cinnamoyl-CoA reductase-like) is quite patchy in streptophyte algae. The diversity of this enzyme family increased independently in several cases in land plants (Figure 4, **Publication V**). The LCA of land plants likely had two homologs of CCR/CCRLs. The evolution of CAD on the other hand, is much clearer: CAD homologs are present throughout the green lineage and the common ancestor of Zygnematophyceae and land plants likely possessed two CAD-like genes. In land plant the CAD family then underwent lineage-specific radiation.

This study discussed the evolutionary history of genes coding for enzyme families that act in early steps of the phenylpropanoid pathway.

For most of the analyzed enzymes, phylogenies hint back at the presence of these enzymes in the last common ancestor of land plants (Figure 10, **Publication V**) or even in the last common ancestor of streptophytes. Furthermore, the performed comparative phylogenomics suggest radiations occurred in several of the enzyme families in an independent lineage-specific manner. To further elucidate the purpose of these genes and if the corresponding enzymes indeed hold similar roles like in land plant phenylpropanoid metabolism has to be resolved with functional studies (see Outlook).

5.2.2 Chromosome-scale genome reveal co-option networks connected to programs for multicellularity (Preprint VII)

Streptophyte algae are organisms that we just begin to understand. In contrast the situation in land plants, where model organisms exist, including the angiosperm *Arabidopsis thaliana*, the moss *Physcomitrium patens*, and the liverwort *Marchantia polymorpha*, there are no fully established model organisms in the grade of streptophyte algae even though as sisters to land plants they are of special interest regarding evolutionary questions. Therefore, the study behind **preprint VII** aimed to give more insight into an emerging model system of Zygnematophyceae, the filamentous alga *Zygnema circumcarinatum*.

Comparative genomic analyses were performed by generating *de novo* genome data on different strains of filamentous algae from the genus *Zygnema*: three strains of the species *Z. circumcarinatum* (SAG 698-1b, UTEX 1559, UTEX 1560) and one strain, SAG 698-1a, that most likely represents *Z. cylindricum* (see Feng *et al.*, 2021) underlying gene networks for multicellularity as well as cell wall synthesis could be uncovered. This study is also the first to generate chromosome-level genome assemblies for any streptophyte algal species. Here, 20 chromosomes could be assembled in *Z. circumcarinatum* SAG 698-1b via Hi-C. This agrees with cytological staining that captured prophase chromosomes (Figure 1b, **preprint VII**). By analyzing further chromatin conformation data, chromosome assemblies of the other two *Z. circumcarinatum* strains were obtained as well. The nuclear genome of SAG 698-1a hints at polyploidy which is not uncommon in Zygnematophyceae (Allen, 1958), whereas the three nuclear genomes of *Z. circumcarinatum* are, as of today, the smallest sequenced genomes in the streptophyte algal grade—while at the same time having the highest protein coding gene density (Figure 1e, **preprint VII**).

Comparative genomics were used to perform an analysis on significantly enriched conserved orthogroups that are shared by the common ancestors of different clades and grades throughout the green lineage (Figure 2a, **preprint VII**).

Orthogroups in the last common ancestor of Zygnematophyceae and land plants reveal GO-terms related to phytohormones, lipids and glucan (Figure 2c, **preprint VII**).

While conserved orthogroups originating in class of Zygnematophyceae underline the stress-related resilience that certain representatives of *Zygnema* represent (see also Pichrtová *et al.*, 2018; Rippin *et al.*, 2017; Rippin *et al.*, 2019; de Vries *et al.*, 2018). The next paragraphs will give a deeper insight into some important molecular findings unraveled by comparative genomics applied in this study. The reconstruction of the evolution of 38 cell wall-related enzyme families revealed that Zygnematophyceae and Embryophyta share all major enzymes for cell-wall-related polysaccharide biogenesis (Figure 3b, **preprint VII**).

Additionally, a gene expansion for subfamilies for polysaccharide backbone synthesis was found in Zygnematophyceae compared to other streptophyte algal classes. A deeper analysis of cell-wall innovation-related enzymes showed the upregulation of CesA/Csl (CELLULOSE SYNTHASE, CS-LIKE) homologs in *Zygnema* under a variety of environmental conditions. Especially CesA1 and CslL1 showed a significant upregulation under 14 hours diurnal light (L-14), cultivation in liquid medium at 4 °C (LQ_4), desiccation stress at 4 °C (Desi_4) for CesA1, and desiccation stress at 4 °C and heavy metal (cadmium chloride, CdCl₂) stress for CslL1 respectively (Figure 3e, **preprint VII**). The differential gene expression in response to this variety of stressors (19 stress conditions) speaks of a very fine-tuned gene-regulatory system adjusting cell-wall-related components in response to specific stressors. A response of cell wall-related genes to osmotic stress was already confirmed for *Zygnema* strain UTEX 1559 (Fitzek *et al.*, 2019). The primary cell wall can absorb heavy metals like cadmium therefore, the upregulation of core components of the cellulose synthase complex in *Zygnema* seems only reasonable (see also stimulating review by Parrotta *et al.*, 2015). Intriguingly it was CslL1 that showed the most upregulation upon heavy metal stress. In the light of this study new Csl families were identified with CslL being one of them, specific only to Zygnematophyceae. Therefore, the stress response here seems to be very specific to this algal class, as land plants cope with this stress by upregulating other components of the cellulose synthase complex (Chen *et al.*, 2019). Lastly, a co-expression analysis (Figure 3c, **preprint VII**) revealed the co-expression with other core genes of the cellulose synthase complex, supporting earlier analyses (Qiao *et al.*, 2021) and revealing an evolutionary conservation of these co-expression networks.

The inference of gene co-expression networks from the RNA-seq data of the *Zygnema* cultures placed in the 19 environmental conditions revealed functional gene modules. Homologs of genes related to (i) cell division and development, (ii) multicellularity, (iii) stress response, (iv) transporters, (v) phytohormones were searched and out of the 406 obtained clusters 150 showed co-occurrence in at least two of these five categories.

This thesis will briefly focus on the findings of related genes belonging to the categories (iii) stress response and (v) phytohormones. The high co-occurrence of the identified clusters in the five different categories hints at a highly connected signaling network in *Zygnema*. Corresponding to this is also the concept of the plant perceptron (Scheres and van der Putten 2017). This concept describes plant biology as a molecular information-processing network of genes. These genes are interwoven on multiple levels. The biochemical interactions that they regulate which respond to changing input cues adequately but also modulate each other, reflecting e.g. the cross-talk between phytohormones.

A highly connected network, as conceptually described in the plant perceptron, can lay the foundation of a biological system that has multicellular characteristics (like filamentous *Zygnema*) and, therefore must act on changing environmental conditions accordingly. Gene homologs for hallmark land plant stress response, including phytohormone signaling, were already present before the origin of land plants (see also **Publication I**, **Publication IV**, **Publication VI**). A broad analysis of phytohormone biogenesis and signaling-related genes confirms their deep evolutionary roots across the green lineage (Figure 5, **preprint VII**). An intriguing example for a land plant phytohormone signaling pathway in streptophyte algae is the ABA-signaling cascade. This cascade along with the other water-related land-plant innovations is well studied (**Publication VI**, see also Bowles *et al.*, 2022). The underlying cascade for ABA-signaling is already present in streptophyte algae (see also overview in **Publication IV**), but the inhibition of downstream PP2C-phosphatase by the PYL-receptor occurs in an ABA-independent manner (Sun *et al.*, 2019). Homologous genes for the underlying chassis for ABA signaling could be recovered in all for *Zygnema* strains which conforms with previous studies on *Zygnema* and Zygnematophyceae respectively (de Vries *et al.*, 2018, Cheng *et al.*, 2019); interestingly, the PP2C were found among the expanded gene family in *Zygnema*. Additionally, ABA could be detected in *Zygnema* by LS-MS even though two ABA biosynthesis, coding for key enzymes in *A.thaliana*, were missing (Figure 5, **preprint VII**), which suggests an alternative ABA synthesis route in *Zygnema*.

Other important findings of the study that were also of relevance to this thesis include the screening of phenylpropanoid enzyme homologs in the *Zygnema* strains. Indeed, previous analyses could be confirmed for the enzymes like 4CL, CAD, and OMT (Figure 5, Figure S14, preprint VIII). The co-expression analysis revealed the presence of stress-responsive homologs like ELIP (early light-inducible proteins), which are upregulated in high light stress as well as dehydration stress (Bartels *et al.*, 1992; Pötter and Kloppstech, 1993) and were present in various clusters (Figure 4b, 4g, **preprint VII**), which speaks of their significance in different stress-responsive signaling pathways.

ELIP homologs were also found in the significantly expanded orthogroups of the common ancestor of all four *Zygnema* strains (Table S3b, **preprint VII**). Interestingly, co-expression analysis also revealed a highly connected homolog coding for lipid-droplet-associated protein Oleosin (*Zci* 13615.1 OLEOSIN) (Figure 4b, **preprint VII**). The expression of OLEOSIN was here observed under osmotic and cold stress conditions, which is reasonable for the LD-related response to various stress conditions. It fits with other studies (**Publication III**) but deserves further insight.

All in all, the comparative analyses in this study unveil deep land-plant-like signaling cascades in *Zygnema* and provide an insight into evolutionary conserved traits. *Zygnema* remains an intriguing model organism to study terrestrialization related questions.

5.2.3 Comparative analyses unravel diversity in Chlorokybophyceae (Publication VIII)

Investigating diversity among the six streptophyte algal classes is a key factor to understand conserved functions aiding plant terrestrialization. In this regard recent advances have been made on the closest algal relatives to land plants, the Zygnematophyceae. This class is the most diverse class with at least 45 genera and over 4000 described species (Guiry and Guiry, 2021), also displaying different morphologies from filamentous to unicellular. This diversity makes the investigation of conserved traits throughout streptophyte evolution not an easy task. Now, Hess and colleagues established a five-order-system for the Zygnematophyceae using comprehensive phylogenomic analyses (Hess *et al.*, 2022). While there are noteworthy advances in the field of Zygnematophyceae trait evolution, many of which have been studied in this thesis, other streptophyte alga classes also deserve further investigation, as these are key lineages for macroevolutionary studies on the streptophyte ancestor. One of those classes will be the focus of discussion in the following paragraph.

Compared to the diversity of the Zygnematophyceae, the diversity of the two most distant streptophyte algae classes to land plants, the Chlorokybophyceae and their sister class (see Lemieux *et al.*, 2007), the Mesostigmatophyceae, was barely investigated. Although genomes for representatives of these two classes were already published (Wang *et al.*, 2020), there was only one genus for both classes, each with two described species for the Mesostigmatophyceae and one described species for the Chlorokybophyceae. This study (**Publication VIII**) now revealed four new species belonging to the Chlorokybophyceae which could be recovered through phylogenomic comparative analyses, adding together with *Chlorokybus atmophyticus*, the genome-sequenced strain, to a total of five species for this streptophyte alga class. *C. atmophyticus* was described by Geitler in 1942 and dwells on terrestrial surfaces like rock surfaces and cracks and is probably cosmopolitan (Geitler, 1942; Rieth, 1972).

Even though *Chlorokybus* is a rare species, eleven isolates could be recovered from all around the world for this study. The construction of a phylogenetic tree with all available strains based on SSU and ITS rDNA analysis (SSU: small subunit, ITS: internal transcribed spacer), two commonly nuclear markers, confirmed deep genetic divergences, even though no apparent morphological differences were found among the eleven isolates (Figure 1, **Publication VIII**).

To validate and gain a deeper insight into these genetic differences, an additional large phylotranscriptomic approach was pursued. RNA was extracted from the isolates, which were all cultured under the same conditions prior to extraction. From the sequenced RNA, *de novo* assemblies of transcriptomes were generated, a set of 529 genes was defined and used for reconstructing a phylogenomic tree; this tree confirmed the deep genetic differences of the *Chlorokybus* isolates (Figure 3a, **Publication VIII**). Interestingly a genetic distance analysis of the isolates turned out to be more than 2-fold larger than the genetic differences among *Arabidopsis* species. Indeed, the Bayesian relaxed molecular clock analysis showed the divergence between *Chlorokybus* could be approx. 76 million years old while the divergence between the *Arabidopsis* species is only between 13 and 28 million years old. Differential gene expression analyses additionally confirmed that there are differences between the isolates (Figure 3b and 3c, **Publication VIII**). Here, protein-coding genes showed high differences among the eight analyzed strains. The newly described *C. riethii* NIES-160 showed the most distinct gene expression profile followed by *C. bremeri* SAG 2611. This could hint at not only deep genetic divergence but at an underlying divergent molecular physiology as well. To scrutinize this notion, further differential gene expression analyses, also in combination with large-scale data from stress-treated isolates, must be performed in the future.

5.2.4 Concluding remarks on Chapter II

Comparative genomics allow deeper insights into conserved underlying molecular mechanisms across the green lineage. This approach sheds light on specific stress-relevant evolutionary conserved gene families by performing phylogenetic screens on a large group of organisms. Comparative studies helped to establish a new algal model organism with intriguing signaling networks and test and formulate new hypotheses regarding plant diversity. Therefore, comparative genomics remain an inherent part of modern research regarding evolutionary questions, especially regarding plant terrestrialization.

6 Outlook

The findings of this thesis contributed to our understanding of streptophyte algal stress response, diversity, and the role of this fascinating group of algae in scientific research questions relating to plant terrestrialization. Yet, many of the topics touched upon herein deserve further attention. Therefore, the following paragraphs will give a brief insight into examples of promising future research topics in the field of streptophyte algal stress response.

In-depth analyses of co-expression networks brought forth the central role of plastid-derived signaling in representatives of Zygnematophyceae (**Publication III and VII**). One of the most connected hubs in the weighted gene co-expression network performed in **Publication III** was a homolog of GOLDEN-LIKE1 (GLK1), a transcription factor that is deeply involved in plastid-to-nucleus communication and is hypothesized to coregulate the expression of nuclear photosynthetic genes in *Arabidopsis*, thereby fine-tuning photosynthetic processes upon a change in environmental conditions (Waters *et al.*, 2009). GLK1 homologs in Zygnematophyceae showed differential gene expression patterns under a diverse set of environmental stress conditions in *Mesotaenium* (**Publication III**), *Zygnema* (**Preprint VII**) and *Mougeotia* (de Vries *et al.*, 2020). Therefore, elucidating the function of GLK genes in streptophyte algae seems reasonable. Studies performed on *Physcomitrium patens* and *Arabidopsis thaliana* suggest a conserved function of GLK throughout Embryophyta (Yasumura *et al.*, 2005). Whether this conservation reaches further back and was already present in the earliest land plants remains to be investigated.

Representatives of the six classes of streptophyte algae already possess the genetic repertoire for the large enzyme families that orchestrate the specialized compound-bearing phenylpropanoid pathway in land plants (**Publication V**). A powerful step towards scrutinizing these phylogenetic findings would be an untargeted metabolomic approach with promising representatives of the streptophyte algae grade. Published mass spectrometric analyses show that compounds like *p*-coumarate, ferulate and phenylpropanoid-derived flavonoids like Naringenin, Kampferol or Genistein occur in representatives of the Chlorophyta (Goiris *et al.*, 2014). Furthermore, in the Zygnematophyceae *Penium margaritaceum*, flavonoid could be detected through mass spectrometry (Jiao *et al.*, 2020). Indeed, *Penium*, also possesses a homolog of the gene coding for the enzyme leading into the flavonoid pathway, the chalcone synthase (CHS) (Jiao *et al.*, 2020). In other streptophyte algae the distribution of genes coding for enzymes involved in this pathway is still quite patchy (see **Publication IV**, de Vries *et al.*, 2017) and deserves further insight.

Some phenylpropanoid-derived compounds are very hard to detect as their production can, as it is the case for other specialized metabolites, fluctuate with the environmental conditions and/or are only transiently present. Therefore, inducing metabolite production through stress treatment can

greatly help to detect certain, possible then enriched compounds. For this, UV-B radiation would be an apt stress treatment: various studies show that UV induces the production of phenylpropanoid-derived compounds like flavonoids throughout the green lineage (e.g., Clayton *et al.*, 2018; Li *et al.*, 1993; Beggs *et al.*, 1985; Apoorva *et al.*, 2021) as well as the production of phenolic compounds in zygnematophycean algae that act as sunscreens (Aigner *et al.*, 2013; Pichrtová *et al.*, 2013).

Phenylalanine ammonia-lyase is the entry enzyme of the phenylpropanoid pathway. As a promising PAL candidate could be found in *Klebsormidium nitens* (**Publication V**, kfl00104_0290_v1.1, 479), an additional homology screen revealed another promising putative PAL candidate in sister species, *Klebsormidium flaccidum* (Figure 9). A first quick alignment of the *K. flaccidum* gene sequence (in Figure 9 TRINITYDN63905-c0-g5-i1) derived from an available transcriptome (co), showed a similarity to the conserved binding sides of the putative *K. nitens* candidate as well as to *Arabidopsis* and *Physcomitrium* PAL sequences.

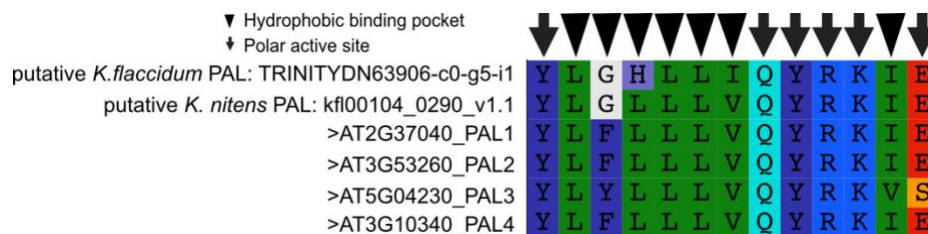


Figure 9: Conserved residues in different putative algae and known *A.thaliana* PAL-candidates. Shown is an alignment of the key residues for substrate binding and function of the phenylalanine ammonia-lyase as reported by Jun *et al.* (2018) and Nagy *et al.* (2019). Next to sequences derived from the alignment file in **Publication V** (Data S1), protein sequence alignments of interest like the *K. nitens* sequence kfl00104_0290_v1.1 as well as the *K. flaccidum* sequence TRINITYDN63905-c0-g5-i1 are shown.

If the promising putative PAL candidates in *K. nitens* and *K. flaccidum* function in the same way as known PALs/PTALs from land plants or possible HALs (meaning binding and converting the substrates L-phenylalanine/L-tyrosine/L-histidine) remains to be investigated by, e.g., heterologous cloning experiments followed by enzyme kinetic analyses.

The genome of *Klebsormidium nitens* already contributed greatly to analyses conducted in this study, especially in **Publication V**. Considering the recent findings this thesis brought forth, as well as the findings from various studies conducted in recent years regarding stress response in Klebsormidiophyceae (e.g., Hartmann *et al.*, 2020; Holzinger and Karsten, 2013), the *de-novo* genome sequencing of *Klebsormidium flaccidum* seems reasonable.

Unraveling the genetic repertoire of this alga might bring forth other promising genes coding for stress responsive enzymes. Another goal in this project is to account for the proper diversity and phylogeny of Klebsormidiophyceae, which is not fully solved as of today—also owing to the broad distribution of Klebsormidiophyceae around the world. Because of this, the sister genus *Interfilum*, which also shows remarkable stress response features (Hartmann *et al.*, 2020; Karsten *et al.*, 2014), is included in this analysis as well. Here, understanding the evolution of diverse morphologies inside the Klebsormidiophyceae through analysis of the underlying genetic repertoire is another overall goal of the study. All in all, adding another *de-novo* genome of a filamentous streptophyte alga, in this case, *Klebsormidium*, can greatly contribute to the understanding of underlying molecular traits facilitating multicellularity throughout streptophyte evolution.

7 References

Adam Z, Rudella A, van Wijk KJ. 2006. Recent advances in the study of Clp, FtsH and other proteases located in chloroplasts. *Current Opinion in Plant Biology* **9**: 234–240.

Aigner S, Remias D, Karsten U, Holzinger A. 2013. Unusual phenolic compounds contribute to ecophysiological performance in the purple-colored green alga *Zygonium ericetorum* (Zygnematophyceae, Streptophyta) from a high-alpine habitat. *Journal of Phycology* **49**: 648–660.

Alboresia A, Gerottob C, Giacomettib GM, Bassia R, Morosinotto T. 2010. *Physcomitrella patens* mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 11128–11133.

Allakhverdiev SI, Kreslavski VD, Klimov VV, Los DA, Carpentier R, Mohanty P. 2008. Heat stress: An overview of molecular responses in photosynthesis. *Photosynthesis Research* **98**: 541–550.

Allen JF. 2015. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proceedings of the National Academy of Sciences* **112**: 201500012–201500012.

Allen MA. 1958. The biology of a species complex in *Spirogyra*. (Ph.D. thesis). *University Bloomington, Indiana, USA*.

Alvarez S, Marsh EL, Schroeder SG, Schachtman DP. 2008. Metabolomic and proteomic changes in the xylem sap of maize under drought. *Plant, Cell and Environment* **31**: 325–340.

Ang LH, Chattopadhyay S, Wei N, Oyama T, Okada K, Batschauer A, Deng XW. 1998. Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of Arabidopsis development. *Molecular Cell* **1**: 213–222.

Apoorva, Jaiswal D, Pandey-Rai S, Agrawal SB. 2021. Untangling the UV-B radiation-induced transcriptional network regulating plant morphogenesis and secondary metabolite production. *Environmental and Experimental Botany* **192**: 104655–104655.

Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. *Current Biology* **25**: R911–R921.

Bailey S, Thompson E, Nixon PJ, Horton P, Mullineaux CW, Robinson C, Mann NH. 2002. A critical role for the Var2 FtsH homologue of *Arabidopsis thaliana* in the photosystem II repair cycle in vivo. *Journal of Biological Chemistry* **277**: 2006–2011.

Baker NR. 2008. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. *Annual Review of Plant Biology* **59**: 89–113.

Bandaranayake WM. 1997. Mycosporines : are they nature 's sunscreens ? : *Nat. Prod. Rep.*, **15**: 159–172.

- Barros J, Dixon RA. 2020.** Plant Phenylalanine/Tyrosine Ammonia-lyases. *Trends in Plant Science* **25**: 66–79.
- Bartels D, Hanke C, Schneider K, Michel D, Salamini F. 1992.** A desiccation-related Elip-like gene from the resurrection plant *Craterostigma plantagineum* is regulated by light and ABA. *EMBO Journal* **11**: 2771–2778.
- Becker B, Marin B. 2009.** Streptophyte algae and the origin of embryophytes. *Annals of Botany* **103**: 999–1004.
- Beggs CJ, Stolzer-Jehle A, Wellmann E. 1985.** Isoflavonoid Formation as an Indicator of UV Stress in Bean (*Phaseolus vulgaris* L.) Leaves. *Plant Physiology* **79**: 630–634.
- Betterle N, Ballottari M, Zorzan S, de Bianchi S, Cazzaniga S, Dall’Osto L, Morosinotto T, Bassi R. 2009.** Light-induced dissociation of an antenna hetero-oligomer is needed for non-photochemical quenching induction. *Journal of Biological Chemistry* **284**: 15255–15266.
- Bittner F, Oreb M, Mendel RR. 2001.** ABA3 Is a Molybdenum Cofactor Sulfurase Required for Activation of Aldehyde Oxidase and Xanthine Dehydrogenase in *Arabidopsis thaliana*. *Journal of Biological Chemistry* **276**: 40381–40384.
- Bonente G, Ballottari M, Truong TB, Morosinotto T, Ahn TK, Fleming GR, Niyogi KK, Bassi R. 2011.** Analysis of LhcSR3, a Protein Essential for Feedback De-Excitation in the Green Alga *Chlamydomonas reinhardtii*. *PLoS Biology* **9**: e1000577–e1000577.
- Borchhardt N, Gründling-Pfaff S. 2020.** Ecophysiological Response Against Temperature in *Klebsormidium* (Streptophyta) Strains Isolated From Biological Soil Crusts of Arctic and Antarctica Indicate Survival During Global Warming. *Frontiers in Ecology and Evolution* **8**: 1–10.
- Borrego SL, Fahrman J, Hou J, Lin DW, Tromberg BJ, Fiehn O, Kaiser P. 2021.** Lipid remodeling in response to methionine stress in MDA-MBA-468 triple-negative breast cancer cells. *Journal of Lipid Research* **62**: 100056–100056.
- Bowles AMC, Paps J, Bechtold U. 2022.** Water-related innovations in land plants evolved by different patterns of gene cooption and novelty. *New Phytol.* **235**(2):732-742.
- Brocard L, Immel F, Coulon D, Esnay N, Tiphile K, Pascal S, Claverol S, Fouillen L, Bessoule JJ, Bréhélin C. 2017.** Proteomic analysis of lipid droplets from *Arabidopsis* aging leaves brings new insight into their biogenesis and functions. *Frontiers in Plant Science* **8**.
- Busch A, Hess S. 2021.** Sunscreen mucilage : a photoprotective adaptation found in terrestrial green algae (Zygnematophyceae). *European Journal of Phycology* **00**: 1–18.
- Buschmann H, Zachgo S. 2016.** The Evolution of Cell Division: From Streptophyte Algae to Land Plants. *Trends in Plant Science* **21**: 872–883.

Breuer G, Lamers PP, Martens DE, Draaisma RB, Wijels RH. 2012. The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresource Technology* **124**: 217–226.

Buschmann H. 2020. Into another dimension: How streptophyte algae gained morphological complexity. *Journal of Experimental Botany* **71**: 3279–3286.

Carella P, Gogleva A, Hoey DJ, Bridgen AJ, Stolze SC, Nakagami H, Schornack S. 2019. Conserved Biochemical Defenses Underpin Host Responses to Oomycete Infection in an Early-Divergent Land Plant Lineage. *Current Biology* **29**: 2282-2294.e5.

Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ. 2016. Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond. *Annual Review of Plant Biology* **67**: 25–53.

Chen H, Li Y, Ma X, Guo L, He Y, Ren Z, Kuang Z, Zhang X, Zhang Z. 2019. Analysis of potential strategies for cadmium stress tolerance revealed by transcriptome analysis of upland cotton. *Scientific Reports* **9**: 1–13.

Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, Castleden I, Song Y, Song B, Huang J, et al. 2016. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant Journal* **85**: 532–547.

Cheng S, Xian W, Fu Y, Melkonian B, Wong GK, Cheng S, Xian W, Fu Y, Marin B, Keller J, et al. 2019. Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell* **179**: 1057–1067.

Chugh M, Reißner M, Bugiel M, Lipka E, Herrmann A, Roy B, Müller S, Schäffer E. 2018. Phragmoplast Orienting Kinesin 2 Is a Weak Motor Switching between Processive and Diffusive Modes. *Biophysical Journal* **115**: 375–385.

Clayton WA, Albert NW, Thrimawithana AH, McGhie TK, Deroles SC, Schwinn KE, Warren BA, McLachlan ARG, Bowman JL, Jordan BR, et al. 2018. UVR8-mediated induction of flavonoid biosynthesis for UVB tolerance is conserved between the liverwort *Marchantia polymorpha* and flowering plants. *Plant Journal* **96**: 503–517.

Cook ME, Graham LE, Archibald JM, Simpson GB, Slamovits CH. 2017. Chlorokybophyceae, Klebsormidiophyceae, Coleochaetophyceae. *Handbook of the protists*. Cham: Springer International Publishing: 1–20.

Copley SD. 2017. Shining a light on enzyme promiscuity. *Current Opinion in Structural Biology* **47**: 167–175.

Correa-Galvis V, Redekop P, Guan K, Griess A, Truong TB, Wakao S, Niyogi KK, Jahns P. 2016. Photosystem II Subunit PsbS Is Involved in the Induction of LHCSR Protein-dependent Energy Dissipation in *Chlamydomonas reinhardtii*. *Journal of Biological Chemistry* **291**: 17478–17487.

Coulon D, Brocard L, Tophile K, Bréhélin C. 2020. *Arabidopsis* LDIP protein locates at a confined area within the lipid droplet surface and favors lipid droplet formation. *Biochimie* **169**: 29–40.

Crawford T, Lehotai N, Strand Å. 2018. The role of retrograde signals during plant stress responses. *Journal of Experimental Botany* **69**: 2783–2795.

Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR. 2010. Abscisic Acid : Emergence of a Core Signaling Network. : 651–682.

Czichi U, Kindl H. 1977. Phenylalanine ammonia lyase and cinnamic acid hydroxylases as assembled consecutive enzymes on microsomal membranes of cucumber cotyledons: Cooperation and subcellular distribution. *Planta* **134**: 133–143.

de Vries J, Archibald JM. 2017. Endosymbiosis: Did Plastids Evolve from a Freshwater Cyanobacterium? *Current Biology* **27**: R103–R105.

de Vries J, Archibald JM. 2018. Plant evolution: landmarks on the path to terrestrial life. *New Phytologist* **217**: 1428–1434.

de Vries J, Curtis BA, Gould SB, Archibald JM. 2018. Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E3471–E3480.

de Vries J, de Vries S, Curtis BA, Zhou H, Penny S, Feussner K, Pinto DM, Steinert M, Cohen AM, von Schwartzberg K, et al. 2020. Heat stress response in the closest algal relatives of land plants reveals conserved stress signaling circuits. *Plant Journal*: 1–24.

de Vries J, de Vries S, Slamovits CH, Rose LE, Archibald JM. 2017. How Embryophytic is the Biosynthesis of Phenylpropanoids and their Derivatives in Streptophyte Algae ? **58**: 934–945.

de Vries J, Gould SB. 2018. The monoplastidic bottleneck in algae and plant evolution. *Journal of Cell Science* **131**: 1–13.

de Vries J, Habicht J, Woehle C, Huang C, Christa G, Wagele H, Nickelsen J, Martin WF, Gould SB. 2013. Is ftsH the Key to Plastid Longevity in Sacoglossan Slugs? *Genome Biology and Evolution* **5**: 2540–2548.

de Vries J, Ischebeck T. 2020. Ties between Stress and Lipid Droplets Pre-date Seeds. *Trends in Plant Science* **25**: 1203–1214.

de Vries J, Stanton A, Archibald JM, Gould SB. 2016. Streptophyte Terrestrialization in Light of Plastid Evolution. *Trends in Plant Science* **21**: 467–476.

Delaux P, Se N, Ane J, Be G. 2013. Evolution of the plant – microbe symbiotic ,toolkit'. *Trends in Plant Science* **18**: 12–18.

- Delosme R, Olive J, Wollman FA. 1996.** Changes in light energy distribution upon state transitions: An in vivo photoacoustic study of the wild type and photosynthesis mutants from *Chlamydomonas reinhardtii*. *Biochimica et Biophysica Acta - Bioenergetics* **1273**: 150–158.
- Delwiche CF, Karol KG, Cimino MT, Sytsma KJ. 2002.** Phylogeny of the genus *Coleochaete* (Coleochaetales, charophyta) and related taxa inferred by analysis of the chloroplast gene *rbcl*. *Journal of Phycology* **38**: 394–403.
- Desaint H, Aoun N, Deslandes L, Vaillau F, Roux F, Berthomé R. 2021.** Fight hard or die trying: when plants face pathogens under heat stress. *New Phytologist* **229**: 712–734.
- Devadasu E, Subramanyam R. 2021.** Enhanced Lipid Production in *Chlamydomonas reinhardtii* Caused by Severe Iron Deficiency. *Frontiers in Plant Science* **12**: 1–12.
- Dixon RA, Paiva NL. 1995.** Stress-induced phenylpropanoid metabolism. *Plant Cell* **7**: 1085–1097.
- Domozych DS, Serfis A, Kiemle SN, Gretz MR. 2007.** The structure and biochemistry of charophycean cell walls: I. Pectins of *Penium margaritaceum*. *Protoplasma* **230**: 99–115.
- Dong NQ, Lin HX. 2021.** Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *Journal of Integrative Plant Biology* **63**: 180–209.
- Dunand C, Crèvecoeur M, Penel C. 2007.** Distribution of superoxide and hydrogen peroxide in *Arabidopsis* root and their influence on root development: Possible interaction with peroxidases. *New Phytologist* **174**: 332–341.
- Emiliani G, Fondi M, Fani R, Gribaldo S. 2009.** A horizontal gene transfer at the origin of phenylpropanoid metabolism: A key adaptation of plants to land. *Biology Direct* **4**.
- Engler A. 1892.** Syllabus der Vorlesungen über Specielle und Medicinisch-Pharmaceutische Botanik: eine Uebersicht über das gesamte Pflanzensystem mit Berücksichtigung der Medicinal-und Nutzpflanzen. *Gebrüder Borntraeger.*: 1-184
- Feng X, Holzinger A, Permann C, Anderson D, Yin Y. 2021.** Characterization of Two *Zygnema* Strains (*Zygnema circumcarinatum* SAG 698-1a and SAG 698-1b) and a Rapid Method to Estimate Nuclear Genome Size of Zygnematophycean Green Algae. *Frontiers in Plant Science* **12**.
- Fitzek E, Orton L, Entwistle S, Grayburn WS, Ausland C, Duvall MR, Yin Y. 2019.** Cell wall enzymes in *Zygnema circumcarinatum* UTEX 1559 respond to osmotic stress in a plant-like fashion. *Frontiers in Plant Science* **10**: 1–15.
- Fučíková K, Hall JD, Johansen JR, Lowe R. 2008.** Desmid flora of the Great Smoky Mountains National Park, USA. *Schweizerbart Science Publisher*.
- Gao JG. 2021.** Tracking the evolutionary innovations of plant terrestrialization. *Gene* **769**: 145203–145203.

Gasulla F, Vom Dorp K, Dombrink I, Zähringer U, Gisch N, Dörmann P, Bartels D. 2013. The role of lipid metabolism in the acquisition of desiccation tolerance in *Craterostigma plantagineum*: A comparative approach. *Plant Journal* **75**: 726–741.

Geitler L. 1942. Neue lultlebige Algen aus Wien. *Österreichische Botanische Zeitschrift* **91**: 49–51.

Gerotto C, Alboresi A, Giacometti GM, Bassi R, Morosinotto T. 2012. Coexistence of plant and algal energy dissipation mechanisms in the moss *Physcomitrella patens*. *New Phytologist* **196**: 763–773.

Gerotto C, Morosinotto T. 2013. Evolution of photoprotection mechanisms upon land colonization: Evidence of PSBS-dependent NPQ in late Streptophyte algae. *Physiologia Plantarum* **149**: 583–598.

Gläßer C, Haberer G, Finkemeier I, Pfannschmidt T, Kleine T, Leister D, Dietz KJ, Häusler RE, Grimm B, Mayer KFX. 2014. Meta-analysis of retrograde signaling in *Arabidopsis thaliana* reveals a core module of genes embedded in complex cellular signaling networks. *Molecular Plant* **7**: 1167–1190.

Goiris K, Muylaert K, Voorspoels S, Noten B, De Paepe D, E Baart GJ, De Cooman L. 2014. Detection of flavonoids in microalgae from different evolutionary lineages. *Journal of Phycology* **50**: 483–492.

Graham LE. 1984. *Coleochaete* and the Origin of Land Plants. *American Journal of Botany* **71**: 603–608.

Grieco M, Jain A, Ebersberger I, Teige M. 2016. An evolutionary view on thylakoid protein phosphorylation uncovers novel phosphorylation hotspots with potential functional implications. *Journal of Experimental Botany* **67**: 3883–3896.

Grolig F, Wagner G. 1988. Light-Dependent Chloroplast Reorientation in *Mougeotia* and *Mesotaenium*: Riased by Pigment-Regulated Plasmalemma Anchorage Sites to Actin Filaments? *Botanica Acta* **101**: 2–6.

Guiry MD, Guiry GM. 2021. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>.

Hall JD, McCourt RM. 2017. Zygnematophyta. *Handbook of the Protists*. Cham, Springer: 135–163.

Hamberger B, Ellis M, Friedmann M, De Azevedo Souza C, Barbazuk B, Douglas CJ. 2007. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: The *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Canadian Journal of Botany* **85**: 1182–1201.

Han X, Chang X, Zhang Z, Chen H, He H, Zhong B, Deng XW. 2019. Origin and Evolution of Core Components Responsible for Monitoring Light Environment Changes during Plant Terrestrialization. *Molecular Plant* **12**: 847–862.

- Hargreaves BR, Girdner SF, Buktenica MW, Collier RW, Urbach E, Larson GL. 2007.** Ultraviolet radiation and bio-optics in Crater Lake, Oregon. *Hydrobiologia* **574**: 107–140.
- Harholt J, Moestrup Ø, Ulvskov P. 2016.** Why Plants Were Terrestrial from the Beginning. *Trends in Plant Science* **21**: 96–101.
- Harris BJ, Clark JW, Schrempf D, Szöllősi GJ, Donoghue PCJ, Hetherington AM, Williams TA. 2022.** Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nature Ecology and Evolution* **6**: 1634–1643.
- Harris BJ, Harrison CJ, Hetherington AM, Williams TA. 2020.** Phylogenomic Evidence for the Monophyly of Bryophytes and the Reductive Evolution of Stomata. *Current Biology* **30**: 2001-2012.e2.
- Hartmann A, Glaser K, Holzinger A, Ganzera M, Karsten U. 2020.** Klebsormidin A and B , Two New UV-Sunscreen Compounds in Green Microalgal *Interfilum* and *Klebsormidium* Species (Streptophyta) From Terrestrial Habitats. **11**: 1–12.
- Heddad M, Adamska I. 2002.** The evolution of light stress proteins in photosynthetic organisms. *Comparative and Functional Genomics* **3**: 504–510.
- Helmchen TA, Bhattacharya D, Melkonian M. 1995.** Analyses of Ribosomal RNA Sequences from Glaucocystophyte Cyanelles Provide New Insights into the Evolutionary Relationships of Plastids. *Journal of Molecular Evolution* **41**: 203-210
- Herburger K, Holzinger A. 2015.** Localization and Quantification of Callose in the Streptophyte Green Algae *Zygnema* and *Klebsormidium*: Correlation with Desiccation Tolerance. *Plant and Cell Physiology* **56**: 2259–2270.
- Hess S, Williams SK, Busch A, Buschmann H, Schwartzberg KV, Hess S, Williams SK, Busch A, Irisarri I, Delwiche CF, et al. 2022.** Report A phylogenomically informed five-order system for the closest relatives of land plants. *Current Biology* **32**: 4473-4482.e7.
- Higashi Y, Okazaki Y, Myouga F, Shinozaki K, Saito K. 2015.** Landscape of the lipidome and transcriptome under heat stress in *Arabidopsis thaliana*. *Scientific Reports* **5**. 10533
- Hoffmann L. 1989.** Algae of Terrestrial Habitats. *Botanical Review* **55**, (2):. 77-105
- Holzinger A, Albert A, Aigner S, Uhl J, Schmitt-kopplin P, Pichrtová M. 2018.** Arctic , Antarctic , and temperate green algae *Zygnema spp* . under UV-B stress : vegetative cells perform better than pre-akinetes. **255**: 1239–1252.
- Holzinger A, Karsten U. 2013.** Desiccation stress and tolerance in green algae: Consequences for ultrastructure, physiological, and molecular mechanisms. *Frontiers in Plant Science* **4**.

Holzinger A, Lütz C, Karsten U. 2011. Desiccation stress causes structural and ultrastructural alterations in the aeroterrestrial green alga *Klebsormidium crenulatum* (Klebsormidiophyceae, Streptophyta) isolated from an alpine soil crust. *Journal of Phycology* **47**: 591–602.

Honkanen S, Small I. 2022. The GENOMES UNCOUPLED1 protein has an ancient, highly conserved role but not in retrograde signalling. *New Phytologist* **236**: 99–113.

Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N, et al. 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications* **5**. 3978

Horváth I, Glatz A, Varvasovszki V, Török Z, Páli T, Balogh G, Kovács E, Nádasdi L, Benkő S, Joó F, et al. 1998. Membrane physical state controls the signaling mechanism of the heat shock response in *Synechocystis* PCC 6803: Identification of hsp17 as a 'fluidity gene'. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 3513–3518.

Hutin C, Nussaume L, Moise N, Moya I, Kloppstech K, Havaux M. 2003. Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4921–4926.

Ischebeck, T, Krawczyk, HE, Mullen, RT, Dyer, JM and Chapman, KD. 2020. Lipid Droplets in Plants and Algae: Distribution, Formation, Turnover and Function. *Seminars in Cell & Developmental Biology* **108** (2020): 82–93.

Jahns P, Holzwarth AR. 2012. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**: 182–193.

James CN, Horn PJ, Case CR, Gidda SK, Zhang D, Mullen RT, Dyer JM, Anderson RGW, Chapman KD. 2010. Disruption of the *Arabidopsis* CGI-58 homologue produces Chananin-Dorfman-like lipid droplet accumulation in plants. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 17833–17838.

Jenkins GI, Long JC, Wade HK, Shenton MR, Bibikova TN. 2001. UV and blue light signalling: Pathways regulating chalcone synthase gene expression in *Arabidopsis*. *New Phytologist* **151**: 121–131.

Jia J, Han D, Gerken HG, Li Y, Sommerfeld M, Hu Q, Xu J. 2015. Molecular mechanisms for photosynthetic carbon partitioning into storage neutral lipids in *Nannochloropsis oceanica* under nitrogen-depletion conditions. *Algal Research* **7**: 66–77.

Jiao C, Sørensen I, Sun X, Sun H, Behar H, Alseekh S, Philippe G, Palacio Lopez K, Sun L, Reed R, et al. 2020. The *Penium margaritaceum* Genome: Hallmarks of the Origins of Land Plants. *Cell* **181**: 1097-1111.e12.

- Jun SY, Sattler SA, Cortez GS, Vermerris W, Sattler SE, Kang CH. 2018.** Biochemical and structural analysis of substrate specificity of a phenylalanine ammonia-lyase. *Plant Physiology* **176**: 1452–1468.
- Kadlubowska J.Z. 1984.** Süßwasserflora von Mitteleuropa. Conjugatophyceae Chlorophyta VIII . Zygneales' **Band 16**, 532 *Gustav Fischer Verlag, Stuttgart*
- Karol KG, McCourt RM, Cimino MT, Delwiche CF. 2001.** The closest living relatives of land plants. *Science* **294**: 2351–2353.
- Karsten U, Herburger K, Holzinger A. 2014.** Dehydration, temperature, and light tolerance in members of the aeroterrestrial green algal genus *Interfilum* (Streptophyta) from biogeographically different temperate soils. *Journal of Phycology* **50**: 804–816.
- Karsten U, Holzinger A. 2012.** Light, Temperature, and Desiccation Effects on Photosynthetic Activity, and Drought-Induced Ultrastructural Changes in the Green Alga *Klebsormidium dissectum* (Streptophyta) from a High Alpine Soil Crust. *Microbial Ecology* **63**: 51–63.
- Kato Y, Sakamoto W. 2009.** Protein quality control in chloroplasts: A current model of D1 protein degradation in the photosystem II repair cycle. *Journal of Biochemistry* **146**: 463–469.
- Kaur H, Heinzl N, Schöttner M, Baldwin IT, Gális I. 2010.** R2R3-NaMYB8 regulates the accumulation of phenylpropanoid-polyamine conjugates, which are essential for local and systemic defense against insect herbivores in *Nicotiana attenuata*. *Plant Physiology* **152**: 1731–1747.
- Khersonsky O, Tawfik DS. 2010.** Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annual Review of Biochemistry* **79**: 471–505.
- Kleine T, Nägele T, Neuhaus HE, Schmitz-Linneweber C, Fernie AR, Geigenberger P, Grimm B, Kaufmann K, Klipp E, Meurer J, et al. 2021.** Acclimation in plants – the Green Hub consortium. *Plant Journal* **106**: 23–40.
- König S, Feussner K, Kaefer A, Landesfeind M, Thurow C, Karlovsky P, Gatz C, Polle A, Feussner I. 2014.** Soluble phenylpropanoids are involved in the defense response of *Arabidopsis* against *Verticillium longisporum*. *New Phytologist* **202**: 823–837.
- Kotak S, Larkindale J, Lee U, Koskull-Döring P von, Vierling E, Scharf KD. 2007.** Complexity of the heat stress response in plants. *Current Opinion in Plant Biology* **10**: 310–316.
- Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J. 2007.** Signals from Chloroplasts Converge to Regulate Nuclear Gene Expression. *Science* **316**: 715–719.
- Krause, W. 1997.** Charales (Charophyceae). - In: Ettl, H; Gärtner, G; Heynig, H, and Mollenhauer, D. (eds): Süßwasserflora von Mitteleuropa, Gustav Fischer Verlag, Stuttgart – Jena. **18**: - 202 pp

Lecointre, G, and Le Guyader, H. 2006. The Tree of Life: A Phylogenetic Classification. Harvard University Press.

Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Lemieux C, Otis C, Turmel M. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology* **5**.

Lenton TM, Dahl TW, Daines SJ, Mills BJW, Ozaki K, Saltzman MR, Porada P. 2016. Earliest land plants created modern levels of atmospheric oxygen. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 9704–9709.

Lewis LA, McCourt RM. 2004. Green algae and the origin of land plants. *American Journal of Botany* **91**: 1535–1556.

Li J, Ou-Lee TM, Raba R, Amundson RG, Last RL. 1993. *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. *Plant Cell* **5**: 171–179.

Li XP, Björkman O, Shih C, Grossman AR, Rosenquist M, Jansson S, Niyogi KK. 2000. A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**: 391–395.

Liebers M, Gillet F, Israel A, Pounot K, Chambon L, Chieb M, Chevalier F, Ruedas R, Favier A, Gans P, et al. 2020. Nucleo-plastidic PAP 8/ pTAC 6 couples chloroplast formation with photomorphogenesis. *The EMBO Journal* **39**: 1–17.

Lind C, Dreyer I, López-Sanjurjo EJ, Von Meyer K, Ishizaki K, Kohchi T, Lang D, Zhao Y, Kreuzer I, Al-Rasheid KAS, et al. 2015. Stomatal guard cells co-opted an ancient ABA-dependent desiccation survival system to regulate stomatal closure. *Current Biology* **25**: 928–935.

Liu B, Long H, Yan J, Ye L, Zhang Q, Chen H, Gao S, Wang Y, Wang X, Sun S. 2021. A HY5-COL3-COL13 regulatory chain for controlling hypocotyl elongation in *Arabidopsis*. *Plant Cell and Environment* **44**: 130–142.

Lois R, Dietrich A, Hahlbrock K, Schulz W. 1989. A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. *The EMBO journal* **8**: 1641–1648.

Marin B, Melkonian M. 1999. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. *Protist* **150**: 399–417.

Martin G, Leivar P, Ludevid D, Tepperman JM, Quail PH, Monte E. 2016. Phytochrome and retrograde signalling pathways converge to antagonistically regulate a light-induced transcriptional network. *Nature Communications* **7**.

Martin WF, Herrmann RG. 1998. Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why? *Plant Physiology* **118**: 9–17.

Martin WF, Garg S, Zimorski V. 2015. Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20140330–20140330.

Mikhailyuk T, Glaser K, Holzinger A, Karsten U. 2015. Biodiversity of *Klebsormidium* (Streptophyta) from alpine biological soil crusts (Alps, Tyrol, Austria, and Italy). *Journal of Phycology* **51**: 750–767.

Mikhailyuk T, Lukešová A, Glaser K, Holzinger A, Obwegeser S, Nyporko S, Friedl T, Karsten U. 2018. New Taxa of Streptophyte Algae (Streptophyta) from Terrestrial Habitats Revealed Using an Integrative Approach. *Protist* **169**: 406–431.

Mikhailyuk TI, Sluiman HJ, Massalski A, Mudimu O, Demchenko EM, Kondratyuk SY, Friedl T. 2008. New streptophyte green algae from terrestrial habitats and an assessment of the genus *Interfilum* (Klebsormidiophyceae, Streptophyta). *Journal of Phycology* **44**: 1586–1603.

Mittler R. 2002. Oxidative stress, antioxidants and stress tolerance. *Trends in Plant Science* **7**: 405–410.

Mochizuki N, Tanaka R, Tanaka A, Masuda T, Nagatani A. 2008. The steady-state level of Mg-protoporphyrin IX is not a determinant of plastid-to-nucleus signaling in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 15184–15189.

Monte I, Kneeshaw S, Franco-Zorrilla JM, Chini A, Zamarreño AM, García-Mina JM, Solano R. 2020. An Ancient COI1-Independent Function for Reactive Electrophilic Oxylipins in Thermotolerance. *Current Biology* **30**: 962-971.e3.

Montgomery W, Potiszil C, Watson JS, Sephton MA. 2016. Sporopollenin, a Natural Copolymer, is Robust under High Hydrostatic Pressure. *Macromolecular Chemistry and Physics* **217**: 2494–2500.

Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E2274–E2283.

Müller P, Li XP, Niyogi KK. 2001. Non-photochemical quenching. A response to excess light energy. *Plant physiology* **125**: 1558–1566.

Mullineaux CW, Allen JF. 1986. The state 2 transition in the cyanobacterium *Synechococcus* 6301 can be driven by respiratory electron flow into the plastoquinone pool. *FEBS Letters* **205**: 155–160.

Nägeli C. 1849. Gattungen einzelliger Algen, physiologisch und systematisch bearbeitet. Neue Denkschriften der Allg. Schweizerischen Gesellschaft für die Gesamten Naturwissenschaften

1849. *Neue Denkschriften der Allgemeinen Schweizerischen Gesellschaft für die gesamten Naturwissenschaften*. **10**:1-139

Nagy EZA, Tork SD, Lang PA, Filip A, Irimie FD, Poppe L, Toşa MI, Schofield CJ, Brem J, Paizs C, et al. 2019. Mapping the Hydrophobic Substrate Binding Site of Phenylalanine Ammonia-Lyase from *Petroselinum crispum*. *ACS Catalysis* **9**: 8825–8834.

Nelson D, Werck-Reichhart D. 2011. A P450-centric view of plant evolution. *Plant Journal* **66**: 194–211.

Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, et al. 2018. The *Chara* Genome: Secondary Complexity and Implications for Plant Terrestrialization. *Cell* **174**: 448-464.e24.

Niyogi KK, Björkman O, Grossman AR. 1997. The roles of specific xanthophylls in photoprotection. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 14162–14167.

Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. 2017. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science* **22**: 53–65.

Oliva J, Rommel S, Fossdal CG, Hietala AM, Nemesio-Gorriz M, Solheim H, Elfstrand M. 2015. Transcriptional responses of Norway spruce (*Picea abies*) inner sapwood against *Heterobasidion parviporum*. *Tree Physiology* **35**: 1007–1015.

Oravecz A, Baumann A, Máté Z, Brzezinska A, Molinier J, Oakeley EJ, Ádám É, Schäfer E, Nagy F, Ulm R. 2006. CONSTITUTIVELY PHOTOMORPHOGENIC1 is required for the UV-B response in *Arabidopsis*. *Plant Cell* **18**: 1975–1990.

Osterlund MT, Hardtke CS, Wei N, Deng XW. 2000. Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* **405**: 462–466.

Pai AA, Luca F. 2019. Environmental influences on RNA processing: Biochemical, molecular and genetic regulators of cellular response. *Wiley Interdisciplinary Reviews: RNA* **10**.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences* **108**: 13624–13629.

Parrotta L, Guerriero G, Sergeant K, Cai G, Hausman JF. 2015. Target or barrier? The cell wall of early- and later-diverging plants vs cadmium toxicity: Differences in the response mechanisms. *Frontiers in Plant Science* **6**: 1–16.

Pattison DI, Davies MJ. 2006. Actions of ultraviolet light on cellular structures: *Cell Structures, Carcinogens and Genomic Instability*. *EXS* **96**: 131-57

Permann C, Herburger K, Niedermeier M, Felhofer M, Gierlinger N, Holzinger A. 2021. Cell wall characteristics during sexual reproduction of *Mougeotia* sp. (Zygnematophyceae)

revealed by electron microscopy , glycan microarrays and RAMAN spectroscopy. *Protoplasma*. **258**: 1261–1275

Pfannschmidt T. 2003. Chloroplast redox signals: how photosynthesis controls its own genes. *Trends Plant Sci* **8(1)**: 33–41

Pichrtová M, Arc E, Stöggel W, Kranner I, Hájek T, Hackl H, Holzinger A. 2016. Formation of lipid bodies and changes in fatty acid composition upon pre-akinete formation in Arctic and Antarctic *Zygnema* (Zygnematophyceae, Streptophyta) strains. *FEMS Microbiology Ecology* **92**: 1–9.

Pichrtová M, Holzinger A, Kulichová J, Ryšánek D, Šoljaková T, Trumhová K, Nemcova Y. 2018. Molecular and morphological diversity of *Zygnema* and *Zygnemopsis* (Zygnematophyceae, Streptophyta) from Svalbard (High Arctic). *European Journal of Phycology* **53**: 492–508.

Pichrtová M, Remias D, Lewis LA, Holzinger A. 2013. Changes in Phenolic Compounds and Cellular Ultrastructure of Arctic and Antarctic Strains of *Zygnema* (Zygnematophyceae, Streptophyta) after Exposure to Experimentally Enhanced UV to PAR Ratio. *Microbial Ecology* **65**: 68–83.

Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology* **27**: 386–391.

PÖTTER E, KLOPPSTECH K. 1993. Effects of light stress on the expression of early light-inducible proteins in barley. *European Journal of Biochemistry* **214**: 779–786.

Pringsheim M. 1862. XXXIII.—On the pro-embryos of the Charæ. *Annals and Magazine of Natural History* **10**: 321–326.

Proteau PJ, Gerwick WH, Garcia-Pichel F, Castenholz R. 1993. The structure of scytonemin, an ultraviolet sunscreen pigment from the sheaths of cyanobacteria. *Experientia* **49**: 825–829.

Pyc M, Cai Y, Gidda SK, Yurchenko O, Park S, Kretschmar FK, Ischebeck T, Valerius O, Braus GH, Chapman KD, et al. 2017a. *Arabidopsis* lipid droplet-associated protein (LDAP) – interacting protein (LDIP) influences lipid droplet size and neutral lipid homeostasis in both leaves and seeds. *Plant Journal* **92**: 1182–1201.

Pyc M, Cai Y, Greer MS, Yurchenko O, Chapman KD, Dyer JM, Mullen RT. 2017b. Turning Over a New Leaf in Lipid Droplet Biology. *Trends in Plant Science* **22**: 596–609.

Qiao Z, Lampugnani ER, Yan X-F, Khan GA, Saw WG, Hannah P, Qian F, Calabria J, Miao Y, Grüber G, et al. 2021. Structure of *Arabidopsis* CESA3 catalytic domain with its substrate UDP-glucose provides insight into the mechanism of cellulose synthesis. *Proceedings of the National Academy of Sciences* **118**.

Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP. 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of Nucleic Acids* **2010**.

Remias D, Holzinger A, Aigner S, Lütz C. 2012a. Ecophysiology and ultrastructure of *Ancylonema nordenskiöldii* (Zygnematales, Streptophyta), causing brown ice on glaciers in Svalbard (high arctic). *Polar Biology* **35**: 899–908.

Remias D, Schwaiger S, Aigner S, Leya T, Stuppner H, Lütz C. 2012b. Characterization of an UV- and VIS-absorbing, purpurogallin-derived secondary pigment new to algae and highly abundant in *Mesotaenium berggrenii* (Zygnematophyceae, Chlorophyta), an extremophyte living on glaciers. *FEMS Microbiology Ecology* **79**: 638–648.

Renault H, Alber A, Horst NA, Lopes AB, Fich EA, Kriegshauser L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, et al. 2017. A phenol-enriched cuticle is ancestral to lignin evolution in land plants. *Nat Commun* **8**: 14713

Rensing SA. 2018. Great moments in evolution: the conquest of land by plants. *Current Opinion in Plant Biology* **42**: 49–54.

Richter AS, Nägele T, Grimm B, Kaufmann K, Schroda M, Leister D, Kleine T. 2022. Retrograde signaling in plants: A critical review focusing on the GUN pathway and beyond. *Plant communications*: 100511–100511.

Richter K, Haslbeck M, Buchner J. 2010. The Heat Shock Response: Life on the Verge of Death. *Molecular Cell* **40**: 253–266.

Rieth A. 1972. Über *Chlorokybus atmophyticus* Geitler 1942. *Arch Protistenkunde*.

Rippin M, Becker B, Holzinger A. 2017. Enhanced Desiccation Tolerance in Mature Cultures of the Streptophytic Green Alga *Zygnema circumcarinatum* Revealed by Transcriptomics. **58**: 2067–2084.

Rippin M, Pichrtová M, Arc E, Kranner I, Becker B, Holzinger A. 2019. Metatranscriptomic and metabolite profiling reveals vertical heterogeneity within a *Zygnema* green algal mat from Svalbard (High Arctic). *Environmental Microbiology* **21**: 4283–4299.

Rizzini L, Favory J, Cloix C, Faggionato D, Hara AO, Kaiserli E, Baumeister R, Schäfer E, Nagy F, Jenkins GI, et al. 2011. Perception of UV-B by the *Arabidopsis* UVR8 Protein. **332**: 103–107.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Current Biology* **15**: 1325–1330.

Rossini L, Cribb L, Martin DJ, Langdale JA. 2001. The Maize Golden2 Gene Defines a Novel Class of Transcriptional Regulators in Plants. *The Plant Cell* **13**: 1231-1244

Rozema J, Björn LO, Bornman JF, Gaberščik A, Häder DP, Trošt T, Germ M, Klisch M, Gröniger A, Sinha RP, et al. 2002. The role of UV-B radiation in aquatic and terrestrial ecosystems-An experimental and functional analysis of the evolution of UV-absorbing compounds. *Journal of Photochemistry and Photobiology B: Biology* **66**: 2–12.

Sachar M, Anderson KE, Ma X. 2016. Protoporphyrin IX: The good, the bad, and the ugly. *Journal of Pharmacology and Experimental Therapeutics* **356**: 267–275.

Saijo Y, Sullivan JA, Wang H, Yang J, Shen Y, Rubio V, Ma L, Hoecker U, Deng XW. 2003. The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. *Genes and Development* **17**: 2642–2647.

Salama HMH, Al Watban AA, Al-Fughom AT. 2011. Effect of ultraviolet radiation on chlorophyll, carotenoid, protein and proline contents of some annual desert plants. *Saudi Journal of Biological Sciences* **18**: 79–86.

Scheres B, Van Der Putten WH. 2017. The plant perceptron connects environment to development. *Nature* **543**: 337–345.

Schönbohm E, Schönbohm E. 1984. Biophenole: Steuernde Faktoren bei der lichtorientierten Chloroplastenbewegung? *Biochemie und Physiologie der Pflanzen* **179**: 489–505.

Schwede TF, Rétey J, Schulz GE. 1999. Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile. *Biochemistry* **38**: 5355–5361.

Shimizu T, Masuda T. 2021. The role of tetrapyrrole- and gun1-dependent signaling on chloroplast biogenesis. *Plants* **10**: 1–19.

Shinozaki K, Yamaguchi-Shinozaki K. 2003. Molecular Mechanisms of Plant Responses and Tolerance of Drought and Cold Stress. In: *Plant Biotechnology 2002 and Beyond*. Dordrecht: Springer Netherlands, 30–37.

Shockey JM, Fulda MS, Browse J. 2003. *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiology* **132**: 1065–1076.

Sinha RP, Klisch M, Häder DP. 1999. Induction of a mycosporine-like amino acid (MAA) in the rice-field cyanobacterium *Anabaena* sp. by UV irradiation. *Journal of Photochemistry and Photobiology B: Biology* **52**: 59–64.

Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F, O'Neill MA, Fei Z, Rose JKC, Domozych DS, Willats WGT. 2011. The charophycean green algae provide insights into the early origins of plant cell walls. *Plant Journal* **68**: 201–211.

Stamenković M, Bischof K, Hanelt D. 2014. Xanthophyll cycle pool size and composition in several *Cosmarium* strains (Zygnematophyceae, Streptophyta) are related to their geographic distribution patterns. *Protist* **165**: 14–30.

Sun A and Guo FQ. 2016. Chloroplast Retrograde Regulation of Heat Stress Responses in Plants. *Frontiers in Plant Science* **7**.

Sun Y, Harpazi B, Wijerathna-Yapa A, Merilo E, de Vries J, Michaeli D, Gal M, Cuming AC, Kollist H, Mosquana A. 2019. A ligand-independent origin of abscisic acid perception.

Proceedings of the National Academy of Sciences of the United States of America **116**: 24892–24899.

Sun Y, Pri-Tal O, Michaeli D, Mosquna A. 2020. Evolution of Abscisic Acid Signaling Module and Its Perception. *Frontiers in Plant Science* **11**: 1–9.

Susek RE, Ausubel FM, Chory J. 1993. Signal Transduction Mutants of *Arabidopsis* Uncouple Nuclear CAB and RBCS Gene Expression from Chloroplast Development. *Cell* **74**(5): 787–99

Szövényi P, Gunadi A, Li FW. 2021. Charting the genomic landscape of seed-free plants. *Nature Plants* **7**: 554–565.

Telfer A, Brudvig GW, Moore TA, Styring S, Rutherford AW, Fromme P, Aro EM. 2002. What is β -carotene doing in the photosystem II reaction centre? *Philosophical Transactions of the Royal Society B: Biological Sciences* **357**: 1431–1440.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**: 123–135.

Tsukagoshi H, Busch W, Benfey PN. 2010. Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. *Cell* **143**: 606–616.

Tzvetkova-Chevolleau T, Franck F, Alawady AE, Dall'Osto L, Carrière F, Bassi R, Grimm B, Nussaume L, Havaux M. 2007. The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. *Plant Journal* **50**: 795–809.

Valledor L, Furuhashi T, Recuenco-Muñoz L, Wienkoop S, Weckwerth W. 2014. System-level network analysis of nitrogen starvation and recovery in *Chlamydomonas reinhardtii* reveals potential new targets for increased lipid accumulation. *Biotechnology for Biofuels* **7**: 1–17.

Van den Hoek, C, Mann, DG, Jahns, HM. 1995. *Algae: An Introduction to Phycology.* Cambridge University Press, Cambridge. 623 pp

Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjana W. 2012. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* **24**: 3506–3529.

Vogt T. 2010. Phenylpropanoid biosynthesis. *Molecular Plant* **3**: 2–20.

Wang S, Li L, Li H, Sahu SK, Wang H, Xu Y, Xian W, Song B, Liang H, Cheng S, et al. 2020. Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants* **6**.

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009. GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*. *Plant Cell* **21**: 1109–1128.

Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C. 2006. Discovery of a Substrate Selectivity Switch in Tyrosine Ammonia-Lyase, a Member of the Aromatic Amino Acid Lyase Family. *Chemistry and Biology* **13**: 1317–1326.

Weinbaum SA, Gressel J, Reisfeld A, Edelman M. 1979. Characterization of the 32,000 Dalton Chloroplast Membrane Protein: III. Probing Its Biological Function in *Spirodela*. *64*(5): 828–832.

Weng J, Joseph P. 2013. Chemodiversity in *Selaginella* : A reference system for parallel and convergent metabolic evolution in terrestrial plants. *Front Plant Sci.* **4**: 119

Whatley JM, 1981. Chloroplast evolution-ancient and modern. *Ann NY Acad Sci.* **361**: 154-65

Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* **111**: E4859–E4868.

Wodniok S, Brinkmann H, Glöckner G, Heidel AJ, Philippe H, Melkonian M, Becker B. 2011. Origin of land plants: Do conjugating green algae hold the key? *BMC Evolutionary Biology* **11**.

Wolf L, Rizzini L, Stracke R, Ulm R, Rensing SA. 2010. The molecular and physiological responses of *Physcomitrella patens* to Ultraviolet-B Radiation. *Plant Physiology* **153**: 1123–1134.

Wu GZ, Bock R. 2021. GUN control in retrograde signaling: How GENOMES UNCOUPLED proteins adjust nuclear gene expression to plastid biogenesis. *Plant Cell* **33**: 457–474.

Xiong F, Komenda J, Kopecký J, Nedbal L. 1997. Strategies of ultraviolet-B protection in microscopic algae. *Physiologia Plantarum* **100**: 378–388.

Yang Y, Benning C. 2018. Functions of triacylglycerols during plant development and stress. *Current Opinion in Biotechnology* **49**: 191–198.

Yang ZK, Niu YF, Ma YH, Xue J, Zhang MH, Yang WD, Liu JS, Lu SH, Guan Y, Li HY. 2013. Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnology for Biofuels* **6**: 1–14.

Yasumura Y, Moylan EC, Langdale JA. 2005. A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell* **17**: 1894–1907.

Yoshida K, Inoue N, Sonobe S, Shimmen T. 2003. Involvement of microtubules in rhizoid differentiation of *Spirogyra* species. *Protoplasma* **221**: 227–235.

Yoshida K, Shimmen T. 2009. Involvement of actin filaments in rhizoid morphogenesis of *Spirogyra*. *Physiologia Plantarum* **135**: 98–107.

Yoshida T, Mogami J, Yamaguchi-Shinozaki K. 2014. ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Current Opinion in Plant Biology* **21**: 133–139.

Young AJ, Frank HA. 1996. Energy transfer reactions involving carotenoids: Quenching of chlorophyll fluorescence. *Journal of Photochemistry and Photobiology B: Biology* **36**: 3–15.

Zhao C, Wang Y, Chan KX, Marchant DB, Franks PJ, Randall D, Tee EE, Chen G, Ramesh S, Phua SY, et al. 2019. Evolution of chloroplast retrograde signaling facilitates green plant adaptation to land. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 5015–5020.

Zheng B, Halperin T, Hruskova-Heidingsfeldova O, Adam Z, Clarke AK. 2002. Characterization of chloroplast Clp proteins in *Arabidopsis*: Localization, tissue specificity and stress responses. *Physiologia Plantarum* **114**: 92–101.

Zheng X, Yang Y, Al-Babili S. 2021. Exploring the Diversity and Regulation of Apocarotenoid Metabolic Pathways in Plants. *Frontiers in Plant Science* **12**.

Zhuo C, Cai J, Guo Z. 2013. Overexpression of early light-induced protein (ELIP) gene from *Medicago sativa ssp. falcata* Increases tolerance to abiotic stresses. *Agronomy Journal* **105**: 1433–1440.

8 Acknowledgments

First and foremost, I want to thank you Jan. You motivated me to return to academia, follow you to Göttingen and provided me with this amazing project. You helped me grow not only on a scientific level but also as a person. I truly learned a lot from you in the last years and will be forever grateful. I value the time I spent in your department that I also had the fortune of building together with you and Sophie. Things were not always easy through the last three years, but I always knew that eventually everything would turn out alright because we are a good team and can manage things together. Coming to Göttingen and joining your department was truly one of the best decisions of my life and I thank you for that!

I also want to express my gratitude towards the members of my thesis advisory committee, Prof. Dr. Ivo Feußner, Prof. Dr. Jörg Stülke and Dr. Maike Lorenz. First, for being on my TAC and taking the time to listen to my progress. Prof Dr. Feußner I want to thank for providing our department a place to start our research when we were still building up our own laboratory and for all the support throughout these last years. Prof Dr. Stülke, I also want to thank for his generosity by giving our new department a place to stay in the Grisebachstraße.

Maike, I want to thank you for all the support – scientific and mental – throughout the last years. I always felt very welcome at the EPSAG also because of our nice chats in between doors and I will dearly miss not coming there frequently to work in the cellar and culture some *Mesotaenium*. I want to further thank the whole EPSAG for always welcoming me with open arms and for the amazing support throughout the years.

A huge thanks goes to the whole department of Applied Bioinformatics! In the last three years you have become like a second family to me. I always felt at ease with all of you and I really enjoyed all the little parties and BBQs we all did together.

Maike, you supported me through thick and thin, especially in the last months, and I am very grateful to have found not only a great colleague but also an amazing friend. Thank you for everything you did for me and just for being there.

Amra, I want to also thank you. I am happy to have found you as a friend and I will always look back fondly on the time we spent together in the last years and our long conversations in the office!

I want to thank you, Tanya, for all your support over the last years be it in the lab. My ambitious projects would not have worked out without your help, and I am also very much looking forward to future wine evenings with you and Thomas.

Tim, I want to also thank you. We went through crazy times in the last years. I was very happy when I was not the only PhD student in our department anymore and I am grateful for all your help throughout the years!

Armin, I am very grateful for everything we achieved together in the last year, and I look back on those long stressful days in the office with a smile because I think we were a good team together with Jan. Thank you for everything you are truly Armin the Wise!

Sophie, I want to thank you for having an open ear for everything and always giving me good advice. I learned a lot from you, and I always enjoy our little talks.

Iker, even if you moved to a new chapter in your life, you are still dearly missed here. I always appreciated our time together in the office.

I also want to thank all the people in this department I did not mention yet who make my everyday office/lab-life great! And especially René. You were always an excellent support in the lab, and I thank you for your patience with my sometimes very spontaneous requests.

I also want to thank the whole department of Plant Biochemistry. I want to thank Ellen, Benedikt, Phillip, Pauline, Tegan, Kirstin, Sabine and many more. I am very grateful that I was able to spend time in your department and that you welcomed me as a guest with open arms during times when things were not easy for all of us.

Finally, I want to thank my dear friends and my family in Düsseldorf who gave me great support and strength over the last three years. Rabab, I want to thank you for always being there no matter what. I am happy and deeply grateful for our friendship even if we are now 300 kilometers apart.

Lastly, I want to thank my parents. Without you, I would not be the person I am today. I am so very grateful for all the love and support you give me!

Table 2: An overview of *Mougeotia sp.* identifiers with a BLASTp hit for *A. thaliana* Lipid droplet identifiers. Additionally, *Mougeotia sp.* identifiers showed BLASTp hits for Mesotaenium identifiers, for which a good BLASTp hits with LD identifier from *A. thaliana* ($p < 0.05$) could be shown in **Puplication II**.

LD protein	A.t. Gene identifier	M.sp. Single Gene Ids	BLAST hit (p-value), edgR_FDR	M.e. Gene identifier	M.sp. Single Gene Ids	BLAST hit (p-value)
OLE7	AT2G25890	-	-	Me1_v2_0235030	-	-
		-	-	Me1_v2_0235140	-	-
CLO3	At2g33380	-	-	Me1_v2_0054690	-	-
HSD1	AT5G50700	-	-	Me1_v2_0117020	TRINITY_DN15941_c0_g1_i2	1.11E-16
		-	-	Me1_v2_0085960	TRINITY_DN15941_c0_g1_i2	1.64E-24
		-	-	Me1_v2_0181840	TRINITY_DN15941_c0_g1_i2	2.46E-13
		-	-	Me1_v2_0202980	TRINITY_DN16084_c0_g2_i3	9E-12
LDAP3	AT3G05500	-	-	Me1_v2_0056100	TRINITY_DN13329_c0_g1_i1	7.24E-28
LDIP	AT5G16550	-	-	Me1_v2_0179900	-	-
PUX10	AT4G10790	TRINITY_DN16099_c2_g1_i3	0.8139	Me1_v2_0178080	TRINITY_DN16099_c2_g1_i3	1.03E-30
			-	Me1_v2_0162420	TRINITY_DN16099_c2_g1_i3	1.78E-55
			-	Me1_v2_0111840	TRINITY_DN16099_c2_g1_i3	5.81E-10
			-	Me1_v2_0162410	TRINITY_DN16099_c2_g1_i3	1.02E-34
			-	Me1_v2_0018250	TRINITY_DN16099_c2_g1_i3	0.00000766
SMT1	AT5G13710	TRINITY_DN12861_c0_g1_i3	0.9599	Me1_v2_0156570	TRINITY_DN12861_c0_g1_i3	3.86E-09
			-	Me1_v2_0214480	TRINITY_DN12861_c0_g1_i3	4.67E-140
CAS	AT2G07050	TRINITY_DN17680_c0_g1_i7	0.8001	Me1_v2_0139020	TRINITY_DN17680_c0_g1_i7	0
LIDL2	AT1G73920	-	-	Me1_v2_0173300	-	-
LDPS	AT3G19920	-	-	Me1_v2_0156150	-	-
LIME1	AT4G33110	TRINITY_DN12667_c0_g1_i3	0.0924	Me1_v2_0053000	TRINITY_DN12667_c0_g1_i3	3.15E-96
LIME2	AT4G33120	-	-	Me1_v2_0052990	TRINITY_DN12667_c0_g1_i3	2.21E-33
LDAH1	AT1G10740	-	-	Me1_v2_0140080	-	-
		-	-	Me1_v2_0140090	-	-
LDDH2	AT1G19400	-	-	Me1_v2_0207420	-	-

Table 3: An overview over *Mesotaenium* identifiers that showed BLASTp hits for hallmark lipid-droplet associated proteins in *Arabidopsis thaliana*. Best hit p-values are indicated as well as single protein IDs that were detected through proteomics. (Me IDs source: **Publication III**, Figure 6)

LD protein	A.t. Gene identifier	M.e. Single Gene Ids	Blast hit (p-value)	M.e. Single Protein Ids
OLE7	AT2G25890	Me1_v2_0235030	8.09E-08	Me1_v2_0235030.1
		Me1_v2_0235140	3.79E-11	Me1_v2_0235140.1
CLO3	AT2G33380	Me1_v2_0054690	5.85E-85	Me1_v2_0054690.2
HSD1	AT5G50700	Me1_v2_0117020	2.80E-22	Me1_v2_0117020.4
		Me1_v2_0085960	6.65E-68	Me1_v2_0085960.1
		Me1_v2_0181840	4.69E-82	-
		Me1_v2_0202980	3.72E-17	-
LDAP3	AT3G05500	Me1_v2_0056100	1.06E-32	Me1_v2_0056100.1
LDIP	AT5G16550	Me1_v2_0179900	6.54E-08	-
PUX10	AT4G10790	Me1_v2_0178080	1.14E-29	-
		Me1_v2_0162420	5.66E-57	-
		Me1_v2_0111840	7.59E-08	-
		Me1_v2_0162410	4.09E-32	Me1_v2_0162410.1
		Me1_v2_0018250	1.10E-13	-
SMT1	AT5G13710	Me1_v2_0156570	1.81E-08	-
		Me1_v2_0214480	5.69E-132	-
CAS	AT2G07050	Me1_v2_0139020	0	Me1_v2_0139020.1
LIDL2	AT1G73920	Me1_v2_0173300	1.03E-81	-
LDPS	AT3G19920	Me1_v2_0156150	1.52E-40	-
LIME1	AT4G33110	Me1_v2_0053000	4.76E-79	Me1_v2_0053000.1
LIME2	AT4G33120	Me1_v2_0052990	3.33E-32	-
LDAH1	AT1G10740	Me1_v2_0140080	8.77E-116	-
		Me1_v2_0140090	3.78E-139	-
LDDH2	AT1G19400	Me1_v2_0207420	1.87E-63	-

Mesotaenium ID	Hit ID Refseq Viridiplantae	Description Viridiplantae	Nies-2285 [<i>K. nitens</i> SAG 355-1a]	Description Nies-2285	TAIR	Description Tair
Me1_v2_0199850.1	PTQ28030.1	Hypothetical protein MARPO_0176s0014 [Marchantia polymorpha]	AOA1Y1ID96	S-adenosyl-L-methionine-dependent methyltransferases	-	-
Me1_v2_0114060.1	PTQ30724.1	Hypothetical protein MARPO_0120s0007 [Marchantia polymorpha]	GWHPBHAQ010243	Hypothetical Proteins	AT1G07380.2	Encodes a neutral ceramidase that is involved in sphingolipid homeostasis and responses to oxidative stress
Me1_v2_0204200.1	KAH9548326.1	Hypothetical protein CY35_11G082700 [Sphagnum magellanicum]	GWHPBHAQ001084	Hypothetical Proteins	AT1G52280.1	Other Names: ATRABG3D, RAB GTPASE HOMOLOG G3D, RABG3D. Description: RAB GTPase homolog G3D;(source:Araport11)
Me1_v2_0185210.1	PNR51064.1	Hypothetical protein PHYPA_010250 [Physcomitrium patens]	GWHPBHAQ012591	Hypothetical Proteins	AT2G21280.2	A nuclear-encoded, plastid-targeted protein (AtSula) whose overexpression causes severe yet stochastic plastid (shown in chloroplasts and leucoplasts) division defects. The protein does not appear to interact with either AtFtsZ proteins when studied in a yeast two-hybrid system.
Me1_v2_0056100.1	GBG83101.1	Hypothetical protein CBR_g36719 [Chara braunii]	GWHPBHAQ009722	Hypothetical Proteins	AT3G05500.1	LDAP3
Me1_v2_0166810.1	KAG6556138.1	Hypothetical protein Mapa_002079 [Marchantia paleacea]	AOA1Y1HSK5	DNA damage-binding protein	AT4G05420.1	damaged DNA binding protein 1A (DDB1A)
Me1_v2_0139020.1	KAG0504631.1	Hypothetical protein KC19_N011500 [Ceratodon purpureus]	GWHPBHAQ007691	Hypothetical Proteins	AT2G07050.1	CAS
Me1_v2_0144170.1	XP_024374024.1	Uncharacterized protein LOC112281605 isoform X1 [Physcomitrella patens]	AOA1Y1HNW9	Hypothetical Proteins	AT3G19340.1	Aminopeptidase (DUF3754)
Me1_v2_0146940.1	KAH9539032.1	Hypothetical protein CY35_15G038100 [Sphagnum magellanicum]	AOA1Y1I5B0	Hypothetical Proteins	AT5G04830.1	Nuclear transport factor 2 (NTF2) family protein
Me1_v2_0235030.1	KAH9300285.1	Hypothetical protein KI387_011868, partial [Taxus chinensis]	GWHPBHAQ005812	Hypothetical Proteins	AT2G25890.2	OLE7
Me1_v2_0204790.3	KAH8944311.1	Hypothetical protein BDL97_13G104100 [Sphagnum fallax]	GWHPBAVF008490	Hypothetical Proteins	-	-
Me1_v2_0172760.1	KAH7440458.1	Hypothetical protein KP509_04G108500 [Ceratopteris richardii]	GWHPBHAQ001239	Hypothetical Proteins	AT5G50700.1	HSD1
Me1_v2_0115750.1	KAH9570058.1	Hypothetical protein CY35_02G020200 [Sphagnum magellanicum]	GWHPBHAQ007861	Hypothetical Proteins	AT5G16880.1	TOL1, TOM1-Like 1; Encodes a member of the Arabidopsis TOL (TOM1-LIKE) family of ubiquitin binding proteins that acts redundantly in the recognition and further endocytic sorting of a PIN-FORMED (PIN)-type auxin carrier protein at the plasma membrane, modulating dynamic auxin distribution and associated growth responses.
Me1_v2_0113070.1	-	-	-	-	-	-

Table 4: BLASTp hits for *Mesotaenium* identifiers shown in the volcano plot in **Puplication III**, Figure 6d.