FLORIAN LENK

DISTRIBUTION AND ECOLOGICAL CHARACTERISTICS OF MEMBERS OF THE *ROSEOBACTER* GROUP

DISTRIBUTION AND ECOLOGICAL CHARACTERISTICS OF MEMBERS OF THE *ROSEOBACTER* GROUP

Dissertation
for the award of the degree
Doctor rerum naturalium

at the Georg-August-Universität Göttingen
within the doctoral programme Biology
at the Georg-August University School of Science (GAUSS)

submitted by

FLORIAN LENK
from Ilmenau

Göttingen, 24/06/2020

THESIS ADVISORY COMMITTEE:

Prof. Dr. Rolf Daniel, Dept. of Genomic and Applied Microbiology, Faculty of Biology and Psychology

PD Dr. Michael Hoppert, Dept. of General Microbiology, Faculty of Biology and Psychology

MEMBERS OF THE EXAMINATION BOARD:

Prof. Dr. Rolf Daniel (*Referee*)

PD Dr. Michael Hoppert (*Co-referee*)

Prof. Dr. Kai Heimel, Dept. of Molecular Microbiology and Genetics, Faculty of Biology and Psychology

Prof. Dr. Stefanie Pöggeler, Dept. Genetics of Eukaryotic Microorganisms, Faculty of Biology and Psychology

Prof. Dr. Gerhard Braus, Dept. of Molecular Microbiology and Genetics, Faculty of Biology and Psychology

Prof. Dr. Jörg Stülke, Dept. of General Microbiology, Faculty of Biology and Psychology

DATE OF THE ORAL EXAMINATION:

09/07/2020

喫茶去

Drink tea.

ABSTRACT

Marine *Rhodobacteraceae* are collectively referred to as the *Roseobacter* group, and due to their functional versatility and high abundance in diverse marine habitats, serve as model organisms in the study of microbial interactions, evolutionary processes, and metabolic pathways.

Two species within this group, which belong to the *Octadecabacter* genus, are of particular interest, as they were isolated from sea ice at both poles, an extreme habitat with regard to temperature, oxidative stress, and nutrient availability. Initial analyses demonstrated their unique position among roseobacters, since they contained an unusually large number of transposable elements, inferred gene duplications, and genome rearrangements. However, a lack of comparable genome sequences from closely related strains left unclear, to what extent these observations are actual evidence of environmental adaptations, and how their genomic features compare to those of temperate strains in the same genus.

This thesis expanded the pool of available *Octadecabacter* genome sequences, and utilised the additional data to examine general genomic properties and the nature of extreme adaptation in this group. Using phylogenomic methods, the phylogenetic history of the *Octadecabacter*-associated strains was reconstructed and evaluated in the wider context of the *Roseobacter* group. Comparative gene content analyses were applied to illustrate which aspects of cellular metabolism and biochemistry are altered in polar *Octadecabacters*, and were put into an evolutionary perspective utilising a model of functional gene content evolution. In addition, the global distribution of individual subgroups within the *Octadecabacters* was examined by means of a metagenomic mapping approach.

The analyses presented here demonstrate that genomes of polar *Octadecabacters* encode more complex metabolic networks, consistent with a broader spectrum of available nutrients and more diverse microbial interactions in sea ice. Genome flexibility, and evolvability in general, constitute important prerequisites for efficient adaptation to this extreme habitat, and are both more pronounced in polar than in temperate *Octadecabacter* genomes. Detection patterns of *Octadecabacters* in metagenome sequences suggest that the currently available polar isolates are members of a cosmopolitan genus that also features non-polar species.

The presented results add to our knowledge of the nature of extreme adaptation and its potential underlying processes in roseobacters, and are a contribution to our goal of understanding the biogeography of this important marine group.

ACKNOWLEDGMENTS

It is no understatement that I could not have finished this thesis without the support of my friends and family, whom I want to thank at this place. First, I want to deeply thank John Vollmers and Andreas Leimbach. The role which your support has played during these years cannot be overstated. Thank you, Andreas, for promoting an open and cooperative form of science, and for sharing your knowledge and experience so that it may benefit others. And thank you, John, for many fruitful discussions, and for your critical input, without which this thesis simply would not be what it is today. Thank you both for your selflessness, for always coming to the help of others without hesitation, and for the sense of community you create around you!

Speaking of community, I want to thank the group's former Ph.D. students, including, but not limited to, Marvin, Chris, and Sascha, for many cheerful hours and interesting conversations. Sascha should be especially thanked for his competent work and support as local sysadmin, since he helped provide the tools for my scientific work and that of others, (mostly) without complaining. Only when I had to be my own sysadmin did I find out, which nuisances this job can hold ready...

We should never forget that we, intentionally or not, shape the future scientific community through the way we treat young academics. Unfortunately, caring for anyone but oneself is poorly incentivised by our current system. Over the years, I had the pleasure of supervising two students during their bachelor thesis. I am grateful for the opportunity to take part in their education, and hope that I could show both that we, as agents, are not bound by the incentives set by the system, and that compassion is better than complacency. Karin and Annika, I wish you nothing but the best for your future.

I would also like to thank my friends outside of the lab, with whom I sometimes found emotional support, but always found fun.

With all the support I received from my friends, none of it would have mattered were it not for my mother. When, after just over three years, I was left without funding for my thesis, she stepped in and privately funded me for the remainder of its duration. This amount of compassion and endurance, after having single-handedly raised two sons – one of them with a disability – speaks of a stronger character than most other people possess, and is deserving of anyone's respect. Thank you, mom!

I also want to thank the community for providing a framework that makes theses such as mine possible in the first place: thank you to the people, who create and maintain open and free resources teaching

people how to program, analyse data, and so much more; thank you, residents of Stack Overflow, for spending your private time to help solve problems of others; and thank you, members of my thesis committee and examination board, for taking the time to read and evaluate my work!

My final thank-you goes out to my little friend, because he always kept my feet warm during long nights. ♡

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

3-MST	3-mercaptopyruvate sulfurtransferase
α-KG	α-ketoglutarate
AAP	aerobic anoxygenic photosynthesis
Ac-CoA	acetyl-CoA
AFP	antifreeze protein
AMP	adenosine monophosphate
ANI	average nucleotide identity

API	application programming interface
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pair
cAMP	cyclic AMP
CDD	Conserved Domains Database
CDS	coding DNA sequence
CoA	coenzyme A
COG	cluster of orthologous groups
CSP	cold-shock protein
DDH	DNA-DNA hybridisation
DHA	dihydroxyacetone
DHAP	DHA phosphate
DHPS	2,3-dihydroxypropane-1-sulfonate
DNA	desoxyribonucleic acid
DOC	dissolved organic carbon
DOM	dissolved organic matter
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DUF	domain of unknown function
EPS	extracellular polysaccharides
`ete3`	Environment for (phylogenetic) Tree Exploration
FDR	false discovery rate
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GB	GenBank
GGDC	genome-to-genome distance calculator
GPDH	glycerol-3-phosphate dehydrogenase
GS	glutamine synthetase
GTA	gene transfer agent

HFB	high-flux backbone
HMP	hydroxymethyl pyrimidine
IC	internode certainty
IMG	Integrated Microbial Genomes
IS	insertion sequence
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG orthology
LCA	latest common ancestor
LGT	lateral gene transfer
MB	marine broth
Mbp	mega base pairs (10^6 bp)
mg-RAST	RAST for metagenomes
ML	maximum likelihood
MLSA	multilocus sequence analysis
MP	maximum parsimony
MTA	methylthioadenosine
MTHF	methyl-tetrahydrofolate
MTR	methylthioribose
MUSCLE	multiple sequence comparison by log-expectation
NAD	nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NGS	next-generation sequencing
NJ	neighbour-joining
OG	orthologous group
PAMC	Polar and Alpine Microbial Collection
pH	*pondus hydrogenii*
POCP	percentage of conserved proteins
POM	particulate organic matter
PSSM	position-specific scoring matrix

RAST	Rapid Annotation using Subsystems Technology
RF	Robinson-Foulds (distance)
RNA	ribonucleic acid
rpsBLAST	reverse position-specific BLAST
rRNA	ribosomal RNA
SAM	S-adenosylmethionine
SBP	Selenium-binding protein
SD	standard deviation
SIMCO	sea ice microbial community
SOP	standard operating procedure
TC	tree certainty
TCA	tricarboxylic acid
TE	transposable element
TLS	translesion DNA synthesis
TRAP	tripartite ATP-independent periplasmic (transporter)
tRNA	transfer RNA
Usp	universal stress protein
UV	ultraviolet

1 INTRODUCTION

The global ocean covers just under 71 % of Earth's surface. With an estimated volume of $1.332{,}4 \times 10^{21}$ L [1], it constitutes our planet's largest consecutive habitat. Its role in climate regulation and nutrient cycling is well established, yet poorly understood. For example, 86 % and 76 % of total global evaporation and precipitation, respectively, take place over the ocean [2, 3], making it the global water cycle's most important component. Nonetheless, for vast oceanic regions, we still lack sufficient flux data to accurately model the profound influence on biogeochemistry and how they are, in turn, influenced by global climate change [4]. Furthermore, the ocean constitutes a major carbon sink, with an estimated 50 % of all non-fossil organic carbon stored in oceanic environments, mostly in the form of dissolved organic carbon (DOC), and in marine sediments (700×10^9 tons and $1{,}750 \times 10^9$ tons, respectively) [5, 6]. Each year, marine phytoplankton converts 50×10^9 tons of inorganic to organic carbon, accounting for about half of the world's primary production [7]. Yet, we have only recently begun to unravel the complex interactions and interrelations between these diverse carbon pools and to map out the role microbes play in this cycle [8]. Particularly the latter aspect is of general interest, considering that even sea water with severely limited nutrient content contains 10^4 bacterial cells/mL, and that the estimated average cell density in the oceans is tenfold this value [9]. It is now widely recognised that microbiota play a decisive role in shaping the marine and terrestrial environments. In order to understand the system as a whole, we need to understand the complex metabolic networks and mutual interactions between its microbial components.

The recent advent of next-generation sequencing (NGS) methods (e. g. [10, 11]) has given us unprecedented capabilities to address this topic in detail [12]. With the availability of an ever-increasing number of bacterial genome sequences, it became apparent that several genomic features can be linked to specific niches in marine habitats [13–15]. Furthermore, cultivation-independent *meta-omics* techniques enable the detailed mapping of such genomic features to system dynamics on the community level and beyond [12, 16]. The following chapter will briefly summarise our current understanding of the prokaryotes' role in marine ecosystems.

1.1 MICROBIAL LIFE IN THE OCEAN

Considering the total biomass of oceanic microbes, their importance in shaping the biochemistry of marine habitats becomes apparent. As noted above, primary production adds approximately 50×10^9 tons of organic carbon annually to the marine food web, mainly by the action of *Dinoflagellates*, *Diatoms*, and *Cyanobacteria*. Since the only other major carbon input sources, atmospheric deposition and riverine input, contribute an estimated 0.2×10^9 tons C/year each [17], it is obvious that the life of heterotrophic organisms, which require carbon compounds for their growth, is centred around areas of primary production. Photosynthesis takes place at the ocean surface layer, down to depths of ≈100 m where about 1 % of incident light remains, sometimes summarised as the *photic zone*. Overall, relatively warm surface waters of the photic zone constitute only approximately 2 % of the ocean volume, whereas the vast majority, 80 %, is cold deep water [17], which has a highly uniform temperature range of around 0 to 3 °C. Here, only little primary production takes place (in the form of chemoautotrophic carbon fixation [18, 19]) and respiratory processes dominate. Nutrients are mainly gained by remineralising the more labile compounds of sinking particulate organic matter (POM). As a consequence and since the respiratory processes are slow, cold deep water is enriched in nutrients compared to the surface water, where most compounds available through primary production are quickly recycled.

Apart from carbon, microbial growth capacities are mostly dictated by the availability of phosphorus, nitrogen and iron, the latter two of which are usually limiting factors in oligotrophic sea water [20]. Reactive nitrogen input – primarily in the form of oxidised nitrogen species, NO_x – is in the order of 20×10^6 tons/year from rivers and 67×10^6 tons/year from the atmosphere [17, 21]. Notably atmospheric nitrogen deposition has increased drastically through human activity since the beginning of industrialisation [21]. A similar amount, 100×10^6 tons, is deposited annually in the ocean environment through microbial N_2-fixation [21].

As indicated above, the distribution of these nutrients varies throughout different water masses. In tropical waters, high levels of solar irradiation increase the water temperature and, due to evaporation, its salinity. This leads to the formation of a strong *thermocline* and *pycnocline*, which prevent mixing of the surface and its subjacent layers, and thereby stratify the water [17]. Due to the constant export of nutrients through sedimentation of POM, the surface waters constitute a largely nutrient-replete, oligotrophic habitat [17]. At high latitudes, where solar irradiation is weaker, the *thermo-* and *pycnocline* are mostly absent [17]. Here, surface waters are cooled down by the cold air and start to sink, thereby pushing the underlying water to greater depths. These water masses then flow along the sea bed, following the earth's topol-

ogy until they resurface at specific points, e. g. the North Pacific [17]. In combination with warmer, wind-driven surface currents, a global-scale circulation of water results, called the *thermohaline circulation*, which is the most important factor in nutrient transport within the marine environment [17].

Thus, microbiota are main contributors to the influx of nutrients into the marine system, whereas geochemical processes in the form of water currents determine their distribution and thereby set the general living conditions for the microbial residents. These conditions vary regionally. For example, while surface waters are usually oligotrophic, upwelling of cold, nutrient-rich deep water can locally increase the amount of available solutes and dissolved organic matter (DOM). Marine sediments and estuaries present additional habitats, which constitute important sources of some nutrients, and sinks for others [17]. Furthermore, the marine environment also includes more extreme niches such as sea ice, or deep sea vents. This variety of habitats predisposes marine bacteria to a large phylogenetic and functional diversity, encompassing both cosmopolitan groups with a world-wide distribution, as well as highly specialised endemic species.

The two most abundant and ubiquitous bacterial groups are *Prochlorococcus* (Cyanobacteria) and *Pelagibacter ubique* (α-Proteobacteria) [22]. Both are adapted to oligotrophic growth conditions, exhibiting small cell sizes and highly streamlined genomes [22, 23]. In contrast, marine vibrios (γ-Proteobacteria), and members of the *Roseobacter* group (α-Proteobacteria) usually possess larger genomes, and generally follow a more copiotrophic life style. Their representatives often live in close association with primary producers [22, 24]. Roseobacters in particular demonstrate a degree of genomic flexibility that lets them functionally adapt to diverse ecological niches (discussed in more detail in a later section).

Such genomic adaptations and their underlying mechanisms are of great scientific interest, both for understanding evolutionary processes, as well as mapping out the factors shaping marine ecology [12, 25]. This thesis will contribute by investigating the adaptation of a group of organisms to a cold and otherwise extreme habitat.

Since cold habitats are the most expansive in the marine environment and pose specific restrictions for their resident microorganisms, their properties and general microbial adaptations will be elucidated in the following chapter.

1.2 LIFE IN COLD HABITATS

The low temperature in habitats of the cryosphere has a direct effect on several aspects of cellular biochemistry, which cells need to address in order to maintain viability. The first is membrane fluidity: at lower temperatures, the lipid bi-layer is in danger of transitioning from a

liquid-crystalline into a gel phase, which ultimately leads to loss of its function. Cells combat this transition by incorporating shorter, as well as (poly-)unsaturated fatty acids, among other modifications [26]. A high number of fatty acid desaturase genes has therefore been associated with cold adaptation in some bacterial genera [15].

Secondly, reaction rates drop exponentially with decreasing temperature [27], and in order to maintain proper function, enzymes undergo distinct changes in their structure and composition. In general, cells seem to trade enzymatic stability for activity at lower temperatures [27], mainly by reducing intra-molecular interactions in the protein core, and by increasing the solvent interactions at its surface [28]. The concomitant change in amino acid composition overall includes an increase in glycine residues, and a decrease in proline content [28]. Since secondary structure elements are crucial for a protein's tertiary structure and general function, it is thought that their content does not differ drastically from mesophilic to cold-adapted enzymes, although structural data indicate that more destabilising amino acids are incorporated into α-helices of the latter [29].

The mentioned alterations also serve to facilitate correct protein folding at low temperatures. Nonetheless, some organisms also require dedicated cold-adapted chaperones for viability [28].

In addition, cells need to counter the stabilising effect cold has on secondary structures of both desoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules, in order to maintain function of the transcriptional and translational machineries. A universal tactic to that end is the expression of cold-shock proteins (CSPs), which bind single stranded DNA and RNA, and thus suppress the formation of secondary structures [30]. Psychrophilic organisms furthermore incorporate more dihydrouridine into transfer RNA (tRNA) molecules than their mesophilic relatives, thereby increasing their flexibility [31]. Translational efficiency is also maintained by the expression of specific accessory proteins to the ribosome in some bacteria [32, 33].

Since temperatures in habitats of the cryosphere are near or below the freezing point of water, microbes need to prevent ice crystal formation in order to maintain viability. One tactic is to lower the freezing point through the accumulation of solutes and ions [34], which is why most of the isolated and described psychrophilic organisms show similarities to halophilic or halotolerant individuals. Frequently produced compatible solutes include glycerol, glycine-betaine, and trehalose, among others [34, 35]. Moreover, many organisms express antifreeze proteins (AFPs), which control the crystallisation of water and thereby avoid physical damage to the cellular environment [36]. Cold-dwelling microorganisms also often produce extracellular polysaccharides (EPS) in order to decrease the freeze point in the extracellular space [37]. Particularly in sea ice, *Diatoms* produce high amounts of EPS for cryoprotection [38].

A further effect of low temperature is the higher solubility of oxygen, which therefore exposes cells to higher levels of oxidative stress. To counter this, the specific genetic equipment of psychrophiles often contains more oxygen-consuming enzymes [15, 39], as well as functions to detoxify O_2^- radicals like dismutases, catalases, and others [15, 40].

As mentioned in the previous section, deep ocean water is the single most extensive low-temperature habitat. Sea ice, on the other hand, covers a comparatively small surface ($\approx 15 \times 10^6$ to 22×10^6 km² [41]), but still harbours dense and highly diverse microbial communities [42]. Because of its relevance for this thesis, its genesis and characteristics as a microbial habitat will be elaborated in more detail below.

Sea ice builds up when a uniform layer of ice crystals forms undisturbed at the air-water interface, and these crystals grow downward due to the continued extraction of heat from the water below (*congelation ice*) [41]. If strong winds mix the upper water layer, small ice crystals form in the mixed layer, and rise to the surface once mixing stops, where they then grow and form *pancake ice* [41]. In both cases, ions and solutes contained in the freezing sea water are extruded and collect between the ice crystals in *brine channels*, pockets of liquid water in which salinity can reach near-saturation levels. From these channels, brine flows back into the sea until the channels become disconnected, which decreases the overall salinity of the ice [41]. Notably, sea ice microorganisms are known to hinder this flow via the production of EPS [43]. During the summer months, the ice melts at the surface due to increased solar irradiation. The melted water can flow back down into the sea through channels in the ice, and in the process flush out nutrients and the resident microorganisms [44]. Thus, sea ice bacteria encounter osmotic stress in the form of phases of both extremely high and low salinity, and can also face expulsion from their habitat. While oxygen solubility is decreased in sea ice brine due to the high salt concentration [45], both poles show seasonally increased levels of heavy metals [46–48], which constitute sources of significant oxidative stress for microorganisms (e. g. [49]). Ultraviolet (UV) radiation likewise varies seasonally, and can reach high intensities at the poles [50], posing another form of stress.

Nonetheless, sea ice is a highly productive habitat, which constitutes the basis of the local marine food web and harbours multiple trophic levels [41, 42]. The densest microbial populations form at the ice-sea interface, where the temperature remains mostly uniform at around −2 °C [44]. This community is called the sea ice microbial community (SIMCO) and contains *Diatoms* as the main primary producers. The numerically dominant prokaryotic groups are Flavobacteria (mostly the genera *Psychrobacter* and *Polaribacter*), and γ-Proteobacteria (*Marinobacter*, *Glaciecola* and *Colwellia*) [42, 51], which live in close association with the primary producers. A high availability of nutrients for bacte-

ria is reflected in the fact that a comparatively large fraction of bacteria from sea ice is readily cultivable [51].

α-Proteobacteria are often less abundant in sea ice than in the underlying sea water [52, 53]. However, they can make up a significant portion of the SIMCO and benefit their *Diatom* hosts [54]. The most abundant genus are usually *Octadecabacters* [51], which belong to the *Roseobacter* group. Although present in cold and otherwise extreme habitats [55, 56], this group of organisms often dominates mesophilic marine communities [24]. It will be described in the following chapter.

1.3 THE *ROSEOBACTER* GROUP

Almost thirty years ago, Shiba instituted the *Roseobacter* genus within the α-Proteobacteria, whose members were characterised by their ability to produce bacteriochlorophyll under aerobic conditions [57]. Since then, this feature has been termed aerobic anoxygenic photosynthesis (AAP) [58], and a growing number of related bacterial strains has been isolated from a variety of environments, together called the *Roseobacter* group.

In general, all marine *Rhodobacteraceae* should be considered roseobacters [59]. Initially, these were thought to form a monophyletic clade, and earlier reconstructions of this group's phylogeny defined five subclades, which subsequently served as the basis for multiple comparative analyses (e. g. [60, 61]). This view was however updated recently by Simon et al., who showed the roseobacters to be paraphyletic to non-associated *Rhodobacteraceae*, and that a terrestrial life style has evolved multiple times in the *Roseobacter* group phylogeny [59]. It should be stressed that the term roseobacter does not constitute a valid taxonomic unit above the genus level. Simon et al. further re-categorised this group into a set of eleven prevalent clades, based on a larger number of comparison genomes, as well as stricter criteria of phylogenetic robustness [59].

The roseobacter lineage likely emerged ≈250 million years ago, coinciding with the radiation of marine *Dinoflagellates* [62, 63]. Its members are often found in association with eukaryotic hosts, and have thus emerged as model organisms for the study of the relating interactions [64, 65]. Consequently, roseobacters are highly abundant in coastal regions and during algal blooms [66]. However, they also occupy a wide range of other marine habitats [24], and thus, due to their wide occurrence, play an important role in shaping the ocean's biogeochemistry (with a potentially strong role in global sulphur cycling [67, 68]). This is why this group has received increased attention in recent years, and why efforts have been made to isolate and sequence more of its members from diverse environments.

As far as we can tell from currently sequenced representatives, *Roseobacter* group bacteria predominantly seem to maintain comparatively

large and flexible genomes, which provide them the metabolic versatility to occupy diverse niches [69, 70]. This tactic stands in contrast to, e. g. *Prochlorococcus* and *Pelagibacter*, which tend towards small and highly streamlined genomes that are optimised to grow in oligotrophic waters, and which numerically dominate this type of habitat [69].

An important contribution to the roseobacters' genomic flexibility are gene transfer agents (GTAs), small, virus-like particles, which pack and transmit fragments of the host's DNA [71]. Earlier investigations found these agents in the vast majority of roseobacter genomes and demonstrated their activity in vivo [72].

This flexibility makes this group of organisms ideal to study the evolution and regulation of specific physiological and metabolic features, and relate these to the adaptation to new habitats where appropriate. Examples of features which have already been analysed in some detail are the degradation of aromatic compounds [73], flagellum gene clusters [74], AAP [75], or genomic adaptations to life in surface ocean water [13].

Besides their high abundance in coastal regions and during algal blooms, the *Roseobacter* group is also present at a number of more extreme habitats, such as sea ice [76], or Antarctic hypersaline lakes [55]. One genus, which harbours multiple isolates from extreme and cold environments, is the *Octadecabacter* genus. Its representatives therefore lend themselves to study extremophilic and psychrophilic adaptation within the *Roseobacter* group, and the following chapter will give a brief overview of this genus and its associated strains.

1.4 THE *OCTADECABACTER* GENUS

Gosink, Herwig, and Staley isolated the first members from Arctic and Antarctic sea ice samples, and named them *Octadecabacter arcticus* and *O. antarcticus*, respectively [76]. They chose the genus name based on the major fatty acid present in these bacteria, which is octadecenoic acid. These species attracted scientific interest at the time, due to their strictly psychrophilic life style, as well as their high sequence similarity despite the large geographic distance. Consequently, their genomes were sequenced and analysed in order to identify common features, as well as functional differences which might be linked to the respective habitats [61, 77]. However, a lack of genomes suitable for comparison reduced the generality of these findings and it remained unclear, in how far functions related to adaptation to life at the poles really are specific (i. e. exclusive) to polar *Octadecabacter* species [77].

In recent years, a number of *Rhodobacteraceae* were isolated from temperate habitats, which, based on their 16S ribosomal RNA (rRNA) sequences, were assigned to the *Octadecabacter* genus. The first of these described in more detail was *O. jejudonensis*, originating from an estuary of Jeju island, South Korea [78] (see Figure 1.1 for an overview of the

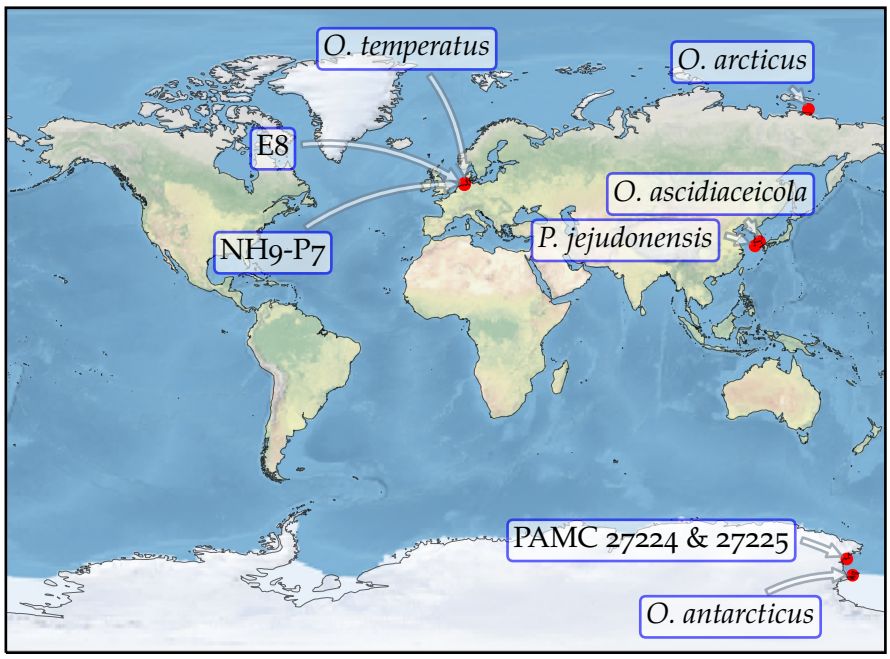


Figure 1.1: The isolation points of *Octadecabacter*-related strains relevant for this thesis.

isolation coordinates of all *Octadecabacter* strains relevant to this study). Other temperate isolates followed soon after, most notably *O. temperatus*, isolated from sea water of the North Sea [79]. In their polyphasic analysis, Billerbeck et al. proposed the reclassification of *O. jejudonensis*, and the institution of a novel genus *Pseudooctadecabacter*, due to its large differences to the polar *Octadecabacters* and temperate *O. temperatus* [79]. Further polar isolates were recently obtained from Antarctic sediment, which were assigned to the *O. antarcticus* species, based on 16S rRNA sequence similarity [80]. Both were deposited at the Polar and Alpine Microbial Collection (PAMC), and initial studies showed that they are psychrotolerant, rather than stenopsychrophiles like the two polar type strains.

Since the polar *Octadecabacters* dwell in one of the most extreme marine-associated habitats known, studying them may prove invaluable for understanding adaptation to extreme environments in the *Roseobacter* group in general. The availability of four polar isolates constitutes a good basis for genomic comparisons and to test the specificity of observed traits. Besides *O. temperatus* and *P. jejudonensis*, further temperate isolates were recently acquired and their genomes sequenced: *Octadecabacter* sp. NH9-P7 was isolated from an Oyster shell, and *Octadecabacter* sp. E8 from sediment, both in the North Sea (Figure 1.1, both unpublished). In addition, the type strain of the new species *O. ascidiaceicola* [81] was genome-sequenced. This strain was isolated from the sea squirt *Halocynthia roretzi* [81].

Taken together, a group of nine closely related strains was available for comparison, of which the genomes of *O. ascidiaceicola*, *P. jejudonensis*, and the PAMC isolates had to be sequenced during this thesis. It should be noted that other cultivated strains exist, which have been classified as *Octadecabacters*, e. g. the recently described *O. ponticola* [82]. The nine strains introduced here (and marked in Figure 1.1) were, however, the only ones available during the time of this thesis' experimental phase, and are consequently the only ones analysed in more detail.

The new isolates' association to the *Octadecabacter* genus relied solely on their 16S rRNA sequences. Such comparisons are, however, associated with great uncertainty [83]. From the analyses of Billerbeck et al. [79], it is already clear that the nine strains described above constitute members of at least two genera. While the definition and delineation of genera is scientifically less than clear [84–86], to avoid confusion, the uncharacterised isolates NH9-P7, E8, and PAMC 27224 and 27225 will only be referred to by their strain identifier from here on, and not assigned to a genus or species. When, throughout the thesis, individual species names are mentioned, these will always refer to the type strains described above, as defined in their original publications (refs. [76, 78, 79]).

In consequence, it is also wrong to speak of this group as the *Octadecabacter* genus, or the *Octadecabacters*. Rather, it is more adequate to use the terms *group* and, if phylogenetically valid, *clade*. The current recommendation for node clade nomenclature is to assign the name of the earliest-described taxon [87]. For informal, e. g. unpublished, clade designations, the name should not be italicised [87]. Consequently, provided that all of the included comparison strains share one common ancestor, which is at the same time exclusive to them, it would be appropriate to speak of the Octadecabacter clade. In the absence of converse information, the term Octadecabacter clade, or simply Octadecabacters, will be used throughout this thesis to refer to this group of nine strains.

1.5 RATIONALE

The majority of habitats on Earth are cold, with the cryosphere covering ≈20 % of its surface [88], and cold deep water making up most of the oceans' volume [17]. Still, only few roseobacter isolates are available from these habitats, although this group is abundant there as well [55, 89, 90]. Two such isolates, *O. arcticus* and *O. antarcticus*, were recently genomically compared [61], and provided first insights into the genetic equipment of cold-adapted roseobacters, albeit without a proper basis for comparison, due to the unavailability of closely related genome sequences.

Since multiple isolates in close relation to the *Octadecabacter* genus have recently become available (some of them from polar habitats),

a more general comparison is possible. It is therefore the aim of this thesis to identify genomic features common to cold-adapted Octadecabacters, using a comparative genomics approach. Comparison with prior analyses should show, in how far mechanisms of cold adaptation known from other organisms are reflected in the polar Octadecabacter genomes. Furthermore, the combined genomic information can provide insights into the source of the polar Octadecabacters' adaptability, which they require to colonise their extreme habitat. As the number of available roseobacter genome sequences has risen steadily in recent years, the general and polar-specific genomic trends observed in Octadecabacters will also be related to other *Roseobacter* group genera and some of their few polar isolates currently available. Notably, comparative analyses within *Roseobacter* group genera concerning adaptation to extreme environments are generally lacking. Therefore, the present thesis is also intended as a reference for future analyses, as more genomic information will no doubt become available over the next years.

This thesis exclusively follows a comparative genomics approach. In a first step, the genomes of four recently isolated, *Octadecabacter*-associated strains (introduced in Section 1.4) will be sequenced using NGS methods, availing a total of nine genomes for comparative analysis. This helps to establish a reliable phylogeny of these genomes in the larger context of the *Roseobacter* group using phylogenomic methods. These methods are better suited than polyphasic studies or comparisons of the 16S rRNA gene for phylogenetic analyses, since they incorporate all of the genomic information available for an organism [91]. Sub-clades in the phylogeny, as well as clusters of increased pairwise sequence similarity will then serve as the basis for gene content comparisons: on the one hand, genome annotation and subsequent orthologue detection are used to determine and interpret orthologous groups (OGs) specific to individual Octadecabacter sub-groups. These observations are complemented with a more general comparison of functional gene content, as captured by the cluster of orthologous groups (COG) classification.

A robust phylogeny furthermore enables to infer changes of e. g. gene content, or sequence characters along ancestral lineages. Phylogenetic birth-and-death models provide a powerful method to that end [92], and were, for example, used to analyse the divergence of life strategies between roseobacters and the closely related SAR 11 lineage [93]. In short, such a model assumes that changes of gene family sizes along the edges of a phylogenetic tree result from a stochastic process, which is characterised by three parameters, κ, λ, and μ. Loss within a family of size n occurs at a rate of $n \times \mu$ and gain at $\kappa + \frac{n}{\lambda}$. All three parameters possess a node-specific and a gene family-specific component, and thus may vary between nodes and families. Given a phylogeny and a corresponding phyletic pattern, the parameters are first optimised on the data in a maximum likelihood (ML) setting. In a second step, the

parameter values can be used to estimate probabilities of gene family changes at the deeper nodes of the phylogeny. This thesis aims to associate specific changes in functional gene content with the colonisation of polar habitats by using a phylogenetic birth-and-death model.

It is important to note that gene content analyses based purely on sequence data have a significant disadvantage: what we can find is always limited by our prior knowledge. With a large number of genomes, one has to rely on automatic annotation to infer gene function, a process which itself can rely on erroneous data (see ref. [94] for one example). Furthermore, research has mostly focused on pathways of the core-metabolism in the past, but these represent only one facet of the cell's biochemical network. As a consequence, we lack a deeper understanding of pathways involved in, e. g. metabolite repair and quality control [95], and a lot of essential genes still possess an unknown, or at best speculative, function [96]. This bias in our current knowledge will inevitably influence the way in which the results of this thesis are interpreted and discussed. While an effort will be made to point out uncertainties and provide alternative explanations where appropriate, it is beyond the scope of this thesis to experimentally test the role and involvement of every component found to differ between polar and temperate strains. Rather, the conclusions drawn here can be the starting point of further experimental analyses concerning the ecology of this exceptional group of organisms.

Finally, beyond the coordinates from which individual Octadecabacters and associated strains were isolated, no investigation into their global distribution and abundance has yet been undertaken. With a multitude of metagenome sequences from a range of sources currently available, a further aim of this thesis is to test and compare the occurrence of sequences related to specific Octadecabacters throughout different habitats and geographic locations.

2 MATERIAL AND METHODS

2.1 GENOME SEQUENCING

Genomes of four strains associated with the *Octadecabacter* genus were sequenced in this study, as noted in Table 2.2. All strains were cultivated in marine broth (MB) medium (Table 2.1) at a temperature of 8 °C (isolates PAMC 27224 and 27225), or 20 °C (*O. ascidiaceicola* and *P. jejudonensis*). Cells were harvested by centrifuging, and their DNA extracted using the MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA), adhering to the manufacturer's instructions. Genomic shotgun paired-end libraries were prepared and sequenced on an Illumina MiSeq (Illumina, San Diego, CA, USA), using the MiSeq reagent kit version 3 according to the manufacturer's instructions.

Table 2.1: Marine broth medium composition. Amounts given refer to one litre of medium. Adjust pH to 7.6 using HCl and NaOH, and sterilise the medium via autoclaving.

COMPONENT	AMOUNT	
Bacto peptone	5.0	g
Bacto yeast extract	1.0	g
Fe(III) citrate	0.1	g
NaCl	19.45	g
$MgCl_2$ (anhydrous)	5.9	g
Na_2SO_4	3.24	g
$CaCl_2$	1.8	g
KCl	0.55	g
$NaHCO_3$	0.16	g
KBr	80	mg
$SrCl_2$	34	mg
H_3BO_3	22	mg
sodium silicate	4	mg
NaF	2.4	mg
$(NH_4)NO_3$	1.6	mg
Na_2HPO_4	8	mg
ddH_2O	ad 1,000	ml

2.2 GENOME ASSEMBLY, ANNOTATION, AND STATISTICS

The sequencing reads obtained from Section 2.1 were quality-filtered using `Trimmomatic`, version 0.32 [97], and subsequently assembled using the `SPAdes` genome assembler, version 3.5.0 [98]. All contigs >500 bp (*O. ascidiaceicola*), or ⩾2,500 bp (*P. jejudonensis*, isolates PAMC 27224 and 27225) were annotated using `Prokka` [99].

Basic genome statistics were summarised through the use of `genomes_feature_table.pl` [100] with the option `-p` to include plasmids in the calculations. Completeness was assessed using `CheckM` [101], which scanned each genome for the presence of 528 marker genes from the *Rhodobacteraceae* family. To assist with examining the presence or absence of specific functions, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were reconstructed in all nine genomes [102]. The annotation of KEGG orthology (KO) terms, as well as the pathway mapping were performed via KEGG's web-interface[1].

2.3 SEQUENCE COMPARISON OF OCTADECABACTER CLADE GENOMES

To clearly delineate species and strain relationships among the Octadecabacters, their genome sequences were subjected to pairwise digital DNA-DNA hybridisation (DDH) [103]. All possible pairings of the nine genomes were submitted to the genome-to-genome distance calculator (GGDC), version 2.1, through the web interface[2] provided by the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ). The recommended formula d_4 [103] and its derived distances were used for downstream analyses.

2.4 BACTERIAL GENOMES USED FOR COMPARATIVE ANALYSIS

A wide range of available genome sequences from members of the *Roseobacter* group were used to create a multilocus sequence analysis (MLSA) phylogeny (Section 2.6), and to analyse their pan-genome (Section 2.10.1). Table 2.2 lists the respective strains, as well as their original publication, where available. Relevant metadata, as well as the genome accession number for each strain are listed in Table A.1 on page 125.

1 https://www.kegg.jp/kegg/mapper.html
2 http://ggdc.dsmz.de/ggdc.php

Table 2.2: Genomes for comparative analysis. The REFERENCE column lists, in order of preference, the publication of a strain's isolation and characterisation, the genome data report, the accession number, or n. a. if none of the former are available; data sources were either GenBank (GB), the Integrated Microbial Genomes (IMG) database, or this study, as indicated in the SOURCE column.

STRAIN	SOURCE	REFERENCE
Aleiiroseovarius crassostreae CV919-312Sm	GB	Boettcher, Barber, and Singer 1999 [104]
Celeribacter baekdonensis B30	GB	AMRK_00000000
Celeribacter indicus P73	GB	Lai et al. 2014 [105]
Citreicella aestuarii 357	GB	Suarez-Suarez et al. 2012 [106]
Citreicella sp. SE45	GB	ACNW_00000000
Rhodobacterales bacterium HTCC2255	GB	NZ_AATR_00000000
Phaeobacter sp. LSS9	IMG	n. a.
Dinoroseobacter shibae DSM 16493	GB	Biebl et al. 2005 [107]
Jannaschia aquimarina GSW-M26	GB	Park and Yoon 2012 [108]
Jannaschia rubra DSM 16279	IMG	Macián et al. 2005 [109]
Jannaschia sp. CCS1	GB	Moran et al. 2007 [70]
Ketogulonicigenium vulgare WSH-001	GB	Liu et al. 2011 [110]
Ketogulonicigenium vulgare Y25	GB	Xiong et al. 2011 [111]
Leisingera aquaemixtae CECT 8399	GB	Park et al. 2017 [112]
Leisingera aquimarina DSM 24565	GB	Vandecandelaere et al. 2008 [113]
Leisingera caerulea DSM 24564	GB	Vandecandelaere et al. 2009 [114]
Leisingera daeponensis DSM 23529	GB	Yoon et al. 2007 [115]
Leisingera methylohalidivorans DSM 14336	GB	Schaefer et al. 2002 [116]
Leisingera sp. ANG1	GB	Collins and Nyholm 2011 [117]

Table 2.2: continued

STRAIN	SOURCE	REFERENCE
Litoreibacter albidus DSM 26922	GB	Romanenko et al. 2011 [118]
Litoreibacter arenae DSM 19593	GB	Kim et al. 2009 [119]
Litoreibacter ascidiaceicola DSM 100566	GB	Kim et al. 2016 [120]
Litoreibacter janthinus DSM 26921	GB	Romanenko et al. 2011 [118]
Loktanella atrilutea DSM 29326	GB	Hosoya and Yokota 2007 [121]
Loktanella cinnabarina LL-001	GB	Tsubouchi et al. 2013 [122]
Loktanella fryxellensis DSM 16213	GB	VanTrappen, Mergaert, and Swings 2004 [55]
Loktanella hongkongensis DSM 17492	GB	Lau et al. 2004 [123]
Loktanella koreensis DSM 17925	GB	Weon et al. 2006 [124]
Loktanella litorea DSM 29433	GB	Yoon, Jung, and Lee 2013 [125]
Loktanella pyoseonensis DSM 21424	GB	Moon et al. 2010 [126]
Loktanella rosea DSM 29591	GB	Ivanova et al. 2005 [127]
Loktanella salsilacus DSM 16199	GB	VanTrappen, Mergaert, and Swings 2004 [55]
Loktanella sediminum DSM 28715	GB	Liang et al. 2015 [128]
Loktanella sp. SE62	IMG	n. a.
Loktanella tamlensis DSM 26879	GB	Lee 2012 [129]
Loktanella vestfoldensis DSM 16212	GB	VanTrappen, Mergaert, and Swings 2004 [55]
Loktanella vestfoldensis SKA53	GB	NZ_AAMS_00000000
Marinovum algicola DG 898	GB	Green et al. 2004 [130]

Table 2.2: continued

STRAIN	SOURCE	REFERENCE
Maritimibacter alkaliphilus HTCC2654	GB	Lee et al. 2007 [131]
Maritimibacter sp. HL-12	GB	NZ_FXBQ_00000000
Nautella italica DSM 26436	GB	Vandecandelaere et al. 2009 [132]
Nautella italica R11	GB	NZ_ABXM_00000000
Nereida ignava CECT 5292	GB	Pujalte et al. 2005 [133]
Oceanibulbus indolifex HEL-45	GB	n. a.
Oceanicola batsensis HTCC2597	GB	Cho and Giovannoni 2004 [134]
Oceanicola granulosus HTCC2516	GB	Cho and Giovannoni 2004 [134]
Oceanicola nanhaiensis DSM 18065	GB	Gu et al. 2007 [135]
Oceanicola sp. HL-35	GB	NZ_JAFT_00000000
Oceanicola sp. MCTG156(1a)	GB	Gutierrez et al. 2017 [136]
Oceanicola sp. S124	GB	Kwon et al. 2012 [137]
Oceaniovalibus guishaninsula JLT2003	GB	Liu et al. 2012 [138]
Octadecabacter antarcticus 307	GB	Gosink, Herwig, and Staley 1997 [76]
Octadecabacter arcticus DSM 13978	GB	Gosink, Herwig, and Staley 1997 [76]
Octadecabacter ascidiaceicola CECT 8868	THIS STUDY	Kim et al. 2016 [81]
Octadecabacter sp. E8	IMG	n. a.
Octadecabacter sp. NH9-P7	IMG	n. a.
Octadecabacter temperatus SB1	GB	Billerbeck et al. 2015 [79]
Octadecabacter sp. PAMC 27224	THIS STUDY	Lee et al. 2014 [80]
Octadecabacter sp. PAMC 27225	THIS STUDY	Lee et al. 2014 [80]
Parvularcula bermudensis HTCC2503	GB	Cho and Giovannoni 2003 [139]
Pelagibaca bermudensis HTCC2601	GB	Cho and Giovannoni 2006 [140]

Table 2.2: continued

STRAIN	SOURCE	REFERENCE
Phaeobacter gallaeciensis BS107	GB	Ruiz-Ponte et al. 1998 [141]
Phaeobacter gallaeciensis DSM 26640	GB	Ruiz-Ponte et al. 1998 [141]
Phaeobacter inhibens DSM 16374	GB	Martens et al. 2006 [142]
Phaeobacter inhibens DSM 17395	GB	Buddruhs et al. 2013 [143]
Planktomarina temperata RCA 23	GB	Giebel et al. 2013 [144]
Pseudophaeobacter arcticus DSM 23566	GB	Zhang et al. 2008 [56]
Pseudooctadecabacter jejudonensis CECT 8397	THIS STUDY	Park and Yoon 2014 [78]
Rhodobacteraceae bacterium SB2	GB	NZ_LGRT_00000000
Rhodobacteraceae bacterium HTCC2083	GB	Kang et al. 2011 [145]
Rhodobacteraceae bacterium HTCC2150	GB	Kang et al. 2010 [146]
Rhodobacterales bacterium Y4I	GB	NZ_ABXF_00000000
Roseobacter denitrificans Och 114	GB	Shiba 1991 [57]
Roseobacter litoralis Och 149	GB	Shiba 1991 [57]
Roseobacter sp. AzwK-3b	GB	Hansel and Francis 2006 [147]
Roseobacter sp. CCS2	GB	NZ_AAYB_00000000
Roseobacter sp. GAI101	GB	NZ_ABXS_00000000
Roseobacter sp. LE17	IMG	n. a.
Roseobacter sp. MED193	GB	NZ_AANB_00000000
Roseobacter sp. R2A57	IMG	n. a.
Roseobacter sp. SK209-2-6	GB	NZ_AAYC_00000000
Roseovarius atlanticus R12B	GB	Li et al. 2016 [148]
Roseovarius mucosus DSM 17069	GB	Biebl et al. 2005 [149]
Roseovarius nubinhibens ISM	GB	González et al. 2003 [150]

Table 2.2: continued

STRAIN	SOURCE	REFERENCE
Roseovarius sp. 217	GB	Schäfer et al. 2005 [151]
Roseovarius sp. MCTG156(2b)	GB	Gutierrez et al. 2017 [152]
Roseovarius sp. TM1035	GB	NZ_ABCL_00000000
Ruegeria atlantica DSM 5823	GB	Rüger and Höfle 1992 [153]
Ruegeria conchae TW15	GB	Lee et al. 2012 [154]
Ruegeria halocynthiae MOLA R1/13b	GB	Doberva et al. 2014 [155]
Ruegeria lacuscaerulensis ITI-1157	GB	NZ_ACNX_00000000
Ruegeria mobilis F1926	GB	Sonnenschein et al. 2017 [156]
Ruegeria pomeroyi DSS-3	GB	González et al. 2003 [150]
Ruegeria sp. TM1040	GB	NZ_AAFG_00000000
Sagittula stellata E-37	GB	Gonzalez et al. 1997 [157]
Sedimentalea nanhaiensis DSM 24252	GB	Sun et al. 2010 [158]
Shimia marina CECT 7688	GB	Choi and Cho 2006 [159]
Shimia sp. SK013	GB	Kanukollu et al. 2016 [160]
Sulfitobacter donghicola DSW-25	GB	Yoon et al. 2007 [161]
Sulfitobacter geojensis MM-124	GB	Kwak et al. 2014 [162]
Sulfitobacter guttiformis KCTC 32187	GB	Labrenz et al. 2000 [163]
Sulfitobacter mediterraneus KCTC 32188	GB	Pukall et al. 1999 [164]
Sulfitobacter noctilucae NB-68	GB	Kwak et al. 2014 [162]
Sulfitobacter noctilucicola NB-77	GB	Kwak et al. 2014 [162]

Table 2.2: continued

STRAIN	SOURCE	REFERENCE
Sulfitobacter pontiacus 3SOLIMAR09	GB	Mas-lladó et al. 2014 [165]
Sulfitobacter pseudonitzschiae H3	GB	Hong et al. 2015 [166]
Sulfitobacter sp. 20_GPM-1509m	GB	NZ_JIBC_00000000
Sulfitobacter sp. CB2047	GB	Ankrah et al. 2014 [167]
Sulfitobacter sp. NAS-14.1	GB	NZ_AALZ_00000000
Sulfitobacter sp. SA11	IMG	n. a.
Tateyamaria sp. ANG1-S1	GB	Collins et al. 2015 [168]
Thalassobium sp. R2A62	GB	NZ_ACOA_00000000
Thalassobacter stenotrophicus CECT 5294	GB	Macián et al. 2005 [169]
Wenxinia marina DSM 24838	GB	Ying et al. 2007 [170]

2.5 ORTHOLOGUE DETECTION

The amino acid sequences of all coding DNA sequences (CDSs) were extracted from the genomes listed in Table 2.2 using `cds_extractor.pl`, version 0.7 [100], and written to multifasta files. `Proteinortho5` [171] was used with the options `-synteny`, `-selfblast` and `-singles` to determine orthologous and paralogous relationships among the extracted proteins. Minimum identity and coverage values for basic local alignment search tool (BLAST) hits to be reported were set to 35 % and 65 %, respectively.

2.6 *ROSEOBACTER* GROUP PHYLOGENY RECONSTRUCTION

core-genome – genes common to all members of a group of organisms

A MLSA phylogeny was reconstructed from the strict, non-redundant (i. e. excluding genes with paralogues) core-genome, of all *Roseobacter* group members listed in Table 2.2. The respective orthologous protein sequences were aligned with the multiple sequence comparison by log-expectation (MUSCLE) algorithm via its homonymous program, version 3.8.31 [172], and the alignments were concatenated to a supermatrix. Positions containing gaps were removed, resulting in a final alignment length of 23,052 patterns (139 genes). Using the `PTHREADS` implementation of `RAxML` (version 8.1.22) [173], the roseobacter phylogeny was reconstructed in a ML-framework under the WAG model of

amino acid substitution (with ML-optimised base frequencies and substitution rates) and a gamma distribution to model rate heterogeneity (discretised into four categories, α parameter estimated through ML). Five hundred bootstraps were computed on random maximum parsimony (MP) starting trees to estimate branch support. Furthermore, ML phylogenies were calculated for each of the 139 core-genes individually, using the same configurations given above. These were then used to estimate bipartition support in the concatenation-based phylogeny through calculation of internode certainty (IC) and tree certainty (TC) scores [174]. In short, bipartitions of the super-matrix-derived phylogeny were evaluated for their frequencies in the 139 individual trees and related to the two most frequent conflicting bipartitions. IC and TC scores reflect this relation, with values close to 0 indicating that multiple conflicting bifurcations occur at the same frequency, and values close to 1 indicating the absence of such conflicts [174]. The scores were calculated using the respective option of `RAxML`.

In addition, a hierarchical clustering was computed for all compared genomes, based on the pairwise Jaccard distances in their gene content. The distances were derived from the orthology information (including singletons), and binarised into presence or absence of individual OGs using the `scikit-learn` package for `python` [175]. Hierarchical clustering was performed using the Farthest Point Algorithm implementation of `scipy` [176] and the clusters were visualised in form of a tree. Two trees were calculated this way, one using the distances derived from all OGs, and one for which core OGs of deeper-branching clades in the phylogeny were marked as absent in all the clade's corresponding members prior to distance calculation. The second tree therefore reflects the similarity of the cloud-genomes of different clades in the phylogeny, as well as the shell-genomes of their individual members. The amount of differing bifurcations between both these trees and the reconstructed phylogeny was expressed as the Robinson-Foulds (RF) distance, and calculated using the Environment for (phylogenetic) Tree Exploration (`ete3`) package for `python` [177].

cloud-genome – genes present in only a subset of all compared genomes

shell-genome – genes present in one or few of the compared genomes

2.7 OCTADECABACTER CLADE PHYLOGENY AND GENE CONTENT CLUSTERING

Individual phylogenies of all 1,513 Octadecabacter core-genes without paralogues were reconstructed in `RAxML`, as described in Section 2.6, with the difference that no bootstrap analyses were performed. The resulting ML trees were combined into a cluster consensus network using `Dendroscope`, version 3.5.9 [178]. Only splits occurring in at least 20 % of trees were visualised. Support of the network's bipartitions was estimated via their frequencies in the 1,513 individual trees and expressed in form of IC and TC scores [174], calculated by `RAxML` (see Section 2.6). In order to compute the scores, two trees were compared

against the rest (reflecting ambiguous positioning of *O. antarcticus*), to ascertain consistency of the calculated scores.

Clustering the Octadecabacters according to their gene content was based on pairwise Jaccard distances, and performed as described in Section 2.6.

2.8 COG ANNOTATION

The COG annotation followed the IMG standard operating procedure (SOP) workflow [179] with a few exceptions. Pre-computed position-specific scoring matrices (PSSMs) of all COGs were acquired from the National Center for Biotechnology Information (NCBI)'s Conserved Domains Database (CDD)[3]. These were used as reference in reverse position-specific BLAST (rpsBLAST) runs with the genomes' protein multifasta files (created in Section 2.5) as queries. Only the best BLAST hit for each protein was reported, with an *e*-value cutoff of 10^{-5}. Tab-separated output was forced using BLAST's `-outfmt 6` option. The script `cdd2cog.pl` version 0.1 [100] parsed the resulting tables and assigned BLAST hits to the appropriate COGs, using the COG descriptions and associations defined in the most recent update from 2014 [180].

2.9 BIRTH-AND-DEATH MODEL OF FUNCTIONAL GENE CONTENT EVOLUTION

A general phylogenetic birth-and-death model of family size evolution was optimised on the COG data, as implemented in the program COUNT [181]. Owing to the model complexity and a lack of parallelisation, optimisation was only computationally feasible for a subset of the 116 strains included in the reference phylogeny. Therefore, a sub-tree of 32 taxa was pruned from the MLSA phylogeny created in Section 2.6, comprising mostly the Octadecabacter and Loktanella clades, more precisely all organisms contained in the equivalents to clades 3, 4, and 5, which Simon et al. recently defined [59]. COG family numbers for each of the 32 genomes were compiled to a phyletic pattern and provided to COUNT, together with the reference phylogeny. Multiple runs of parameter estimation were performed with successively increasing model complexities, starting from uniformity of parameters across tree edges and gene families, and ending with gamma-distributed parameter values, discretised into two to three categories. Complexity of the model was only further increased if the increase in likelihood justified the higher number of free parameters. The final model contained two rate categories for gene loss, and three categories each for duplication, transfer, and length. Using the estimated parameter values, probabilities of gain, loss, expansion, and reduction for each COG at each tree node were calculated by COUNT. For downstream analysis, only those

3 ftp.ncbi.nih.gov/pub/COG/COG2014/data/

COGs were considered, which showed a probability of change >0.5 at the particular node in question.

2.10 GENE CONTENT ANALYSES

2.10.1 *Pan- and core-genome calculations*

pan-genome – union of all genes within a group of organisms

The pan-genome and strict core-genome sizes of the Octadecabacters were both derived from the orthology information obtained in Section 2.5. Their dependence on the number of included genomes was determined by calculating both for each possible combination of one to nine Octadecabacter genomes (without replacement) and subsequent least-squares fit. A power law (Equation (2.1)) and exponential decay (Equation (2.2)) function were fitted, respectively, to the median values of pan- and core-genome size using the `scipy` module for the `python` programming language [176]. In an equivalent analysis for the whole *Roseobacter* group (genomes listed in Table 2.2, excluding the outgroup genome *Parvularcula bermudensis* and *Phaeobacter* sp. LSS9), approximately 600 combinations were drawn at random for each number of species included, as testing all possible combinations was not computationally feasible.

$$f(x) = a \cdot x^b + c \tag{2.1}$$

$$f(x) = a \cdot e^{b \cdot x} + c \tag{2.2}$$

2.10.2 *Determination of group-specific orthologous groups*

The Octadecabacter clade genomes were divided into four groups, according to the clusters based on their pairwise distances in gene content (Section 2.7). Group I contained the polar species *O. arcticus*, *O. antarcticus*, and isolates PAMC 27224 and 27225, group II the temperate species *O. temperatus* and *O. ascidiaceicola*, group III *P. jejudonensis* and isolate NH9-P7 and group IV the single genome of isolate E8. Together with the orthology data, the script `po2group_stats.pl`, version 0.1.3 [100] determined which OGs were specific to each group, i. e. which genes occurred in all members of a group and none of the other compared genomes.

2.11 PROTEIN SECONDARY STRUCTURE PREDICTION AND COMPARISON

The secondary structure content of each annotated protein in the nine Octadecabacter genomes was predicted from the primary amino acid sequence using the algorithm implemented in the `JPred4` secondary structure prediction server [182]. `JPred4` returns the secondary structure state of each amino acid position in the protein (either helix, sheet,

or none), along with a confidence score. To summarise the total content of α-helix and β-sheet, amino acids, which were assigned to one of the two, were counted fully if their confidence score was at least 7, and as one-half if it was 5 or 6. The absolute number of amino acids of each structural feature was divided by the length of the protein to obtain the relative content of both α-helix and β-sheet. Utilising the orthology information (Section 2.5), the nine genomes were hierarchically clustered based on the pairwise euclidean distances in relative secondary structure content of the core-proteome. Clustering was performed as described in Section 2.6.

Statistical difference in secondary structure content was tested separately for helix and sheet through two-sided Wilcoxon signed-rank tests, as implemented in `scipy` [176]. This tests the null-hypothesis H_0 that the median difference between two samples equals zero. For each possible pairing of two Octadecabacter clade genomes, the relative structure content of either helix or sheet in all core proteins served as the observations, based on which the test statistic and the corresponding p-value were computed. In order to assert that the median difference between samples $\neq 0$, i. e. to accept H_1, the significance level α was corrected for multiple statistical testing via the method described by Benjamini and Hochberg [183]. In short, the p-values were sorted in ascending order and assigned ranks. Then, each was compared to its Benjamini-Hochberg critical value given by $\frac{i}{m} \times \alpha$, where i is the value's rank, $m = 36$ the total number of performed tests and $\alpha = 0.05$ the desired significance level, corresponding to the false discovery rate (FDR). The highest p-value, for which $p < \frac{i}{m} \times \alpha$, and all values smaller than it were considered significant in rejecting H_0 and indicating that the respective pair of Octadecabacters shows differences in the particular structural feature. The direction of this difference was tested in subsequent one-sided Wilcoxon signed-rank tests, with the hypothesis H_0 that the median difference between the samples is either positive, or negative.

Benjamini-Hochberg correction controls the false discovery rate

General comparisons of amino acid content were carried out for Octadecabacter strains with highly similar GC-contents, which excluded *P. jejudonensis* and isolate NH9-P7. In χ^2 tests, observed numbers of the twenty amino acids were compared to their expected numbers from the base frequencies (averaged over all seven clade members) for each strain. Mean frequencies of the amino acids in general, as well their ratios of occurrence in α-helices versus non-helix regions, were compared between the genome groups defined in Section 2.10.2.

2.12 OCTADECABACTER OCCURRENCE IN METAGENOMES

The global distribution of Octadecabacters was studied through metagenomic read mapping. The workflow consisted of three main steps:

(i) selection of metagenomes, (ii) read mapping, and (iii) filtering the obtained results for unspecific hits and subsequent normalisation.

2.12.1 *Selection of metagenomes*

In order to keep the workflow computationally feasible, only metagenomes available on the Rapid Annotation using Subsystems Technology (RAST) for metagenomes (mg-RAST) platform [184] were considered. In a pre-selection step, identifiers and meta-data of metagenomes, in which Octadecabacters constituted at least 1 % of reads assigned to the α-Proteobacteria, were acquired. These metagenomes' reads were then downloaded through the mg-RAST application programming interface (API) and used for subsequent mapping.

2.12.2 *Read mapping*

Due to their high conservation, rRNA gene clusters were excised from the Octadecabacter genomes prior to read mapping. Their positions were identified using `rnammer` [185] and the sequence between the coordinates excised. The genome DNA sequences were then concatenated into a single file. `Bowtie2` [186], version 2.3.1, was used to map the reads. Alignment seed length was set to 12, with at maximum one mismatch allowed per seed. Penalties for mismatches, as well as undefined bases ("N"), were set to 3, gap openings and extensions in both reference and read were penalised with 3 and 5, respectively. The minimum score x for a hit to be reported had to satisfy $x \geqslant -0.2 \times \text{readlength} - 6$, corresponding to a minimum nucleotide identity of approximately 90 % between read and reference. Only the best hit was reported.

2.12.3 *Filtering and normalisation*

Unspecific hits were filtered using the available *Roseobacter* group genome information (Table 2.2). To that end, their nucleotide sequences were split into fragments of 1,000 nt length, which were then subjected to pairwise BLAST comparisons between the genomes. Both steps are part of the average nucleotide identity (ANI) comparison workflow of the `python` package `pyANI` [187]. Subsequently, reads which mapped with at least 65 % of their length to Octadecabacter regions with >90 % nucleotide identity to regions in other roseobacters were discarded, and the remaining hits were considered specific to Octadecabacters. The comparison of nucleotide identity included comparisons between Octadecabacter groups, but not within them. This led to the exclusion of less reads, but restricted the analysis to the level of Octadecabacter groups.

As a means of normalisation, for each metagenome the number of nucleotides in specific hits was divided by the number of nucleotides in reads assigned to the bacteria kingdom by mg-RAST.

In order to differentiate between individual Octadecabacters, the same filtering and normalisation procedure described above was applied once more, but included the pairwise comparisons between all nine Octadecabacter genomes.

3 RESULTS

3.1 GENOME ASSEMBLY AND GENERAL COMPARISON

General metrics of the analysed Octadecabacter clade genomes are given in Table 3.1. Despite being highly fragmented with ≈130 contigs, the two PAMC isolate genomes are over 99 % complete, based on the presence of 528 *Rhodobacteraceae* marker genes. Polar Octadecabacters possess, on average, larger genomes than temperate clade members (by ≈800 CDSs), combined with a lower coding percentage. The DNA GC-content is about 55 % in most strains, and markedly higher (about 60 %) in isolate NH9-P7 and *P. jejudonensis*. Based on the orthology data, the Octadecabacter clade's core-genome size is 1,609 (compared to 142 over the whole *Roseobacter* group, including the outgroup genome). Corresponding NCBI accession numbers are provided in Table A.1, p. 125.

Table 3.1: Basic metrics of Octadecabacter clade genomes. STATUS is listed as either finished (F), or (permanent) draft (D). For finished genomes, grey numbers in parentheses indicate the number of plasmids. Plasmids were included in determining genome size, contig number, and CDS count.

	O. antarcticus	*O. arcticus*	PAMC 27224	PAMC 27225	*O. temperatus*	*O. sciadiacecicola*	NH9-P7	E8	*P. jejudonensis*
SIZE [Mbp]	4.875	5.479	4.095	4.124	3.265	3.228	3.549	3.534	3.405
STATUS	F	F	D	D	F	D	D	D	D
CONTIGS	2 (1)	3 (2)	132	129	2 (1)	11	18	13	20
COMPLETENESS	— >99 % —								
GC %	54.62	55.15	55.11	55.15	54.68	54.94	59.88	54.35	59.57
CDSs	4,569	4,694	4,188	4,161	3,294	3,283	3,484	3,484	3,345
CODING %	80.91	78.34	88.77	88.71	91.86	91.51	91.40	90.44	91.22
COGs	3,548	3,851	3,267	3,273	2,717	2,671	2,883	2,914	2,772

Pairwise genome similarities were determined as digital DNA-DNA hybridisation (DDH) values using the genome-to-genome distance calculator (GGDC) [103] and are visualised in Figure 3.1. Notably, the two

PAMC isolates show 98 % similarity to each other, but only ≈30 % to *O. antarcticus*. The genome of *O. ascidiaceicola* is most similar to *O. temperatus*, albeit at a low DDH value of ~25 %, while the two strains E8 and NH9-P7 are less than 20 % similar to any other clade member analysed (the same is true for *Pseudooctadecabacter jejudonensis*). Based on the DDH-derived distances, isolate NH9-P7 and *P. jejudonensis* form a separate cluster, which coincides with their higher GC-content (Table 3.1). The novel polar isolates PAMC 27224 and PAMC 27225 form a distinct and comparatively deep-branching cluster with the type strains of the polar species *O. antarcticus* and *O. arcticus*, neighbouring a corresponding non-polar cluster formed by *O. temperatus* and *O. ascidiaceicola*.

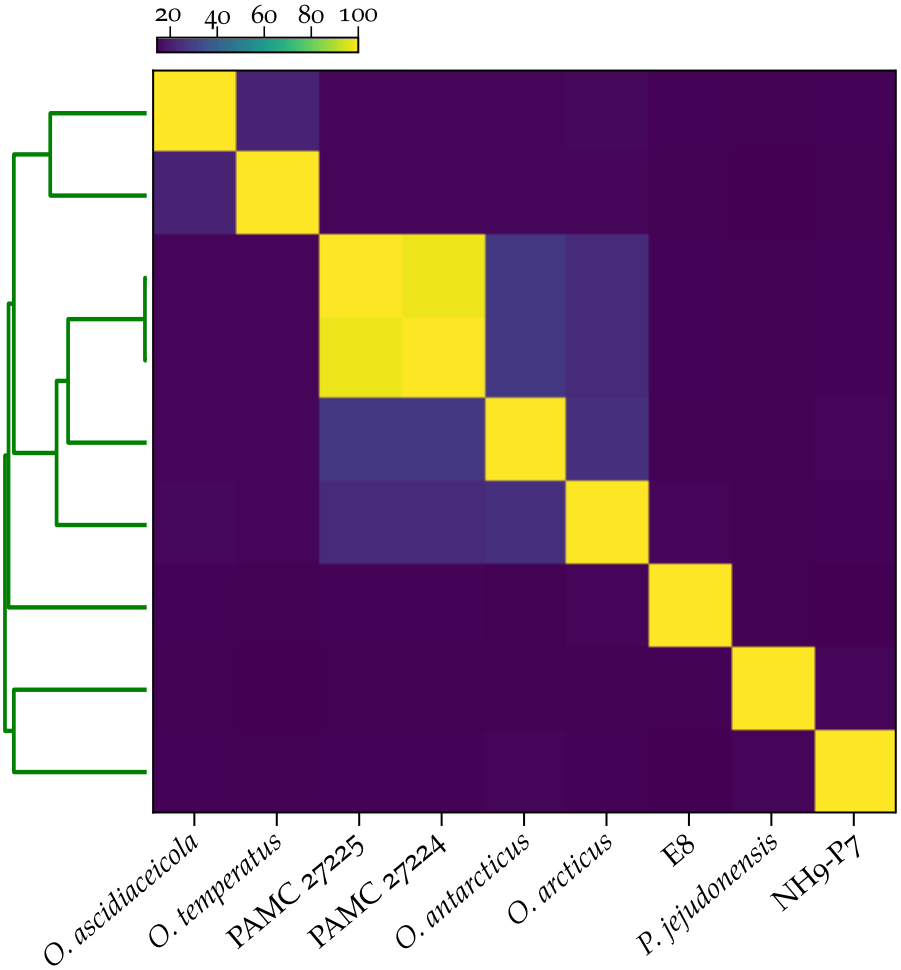


Figure 3.1: Percent similarity values of pairwise Octadecabacter clade genome comparisons via GGDC, presented as a heat-map. Dendrogram on the left represents hierarchical clustering based on the pairwise distances calculated from the DDH values.

3.2 PHYLOGENY

A MLSA-phylogeny based on 139 core-genes of the *Roseobacter* group (including the outgroup genome *Parvularcula bermudensis*) generally displays the same topology, which other studies observed [59, 60] and clusters defined therein can be reproduced here (marked respectively in Figure A.2, p. 131). It also verifies the newly sequenced strains association with the Octadecabacters, since they form a coherent clade with strong bootstrap support (Figure 3.2). Their inferred phylogenetic relationships also mirror the similarity- (DDH-)based observations (see Figure 3.1). The earliest branching and therefore most distant member is isolate NH9-P7, followed by *P. jejudonensis*, although this bifurcation is associated with some uncertainty (low bootstrap support value of 64). All polar isolates cluster together and form a common clade with *O. temperatus* and *O. ascidiaceicola*, which means that they share a more recent common ancestor with each other than with the rest. Within the polar representatives, isolates from the Antarctic form a group distinct from the Arctic species. Isolate E8 is closer related to the temperate/polar clade than to NH9-P7 and *P. jejudonensis*.

Based on the phylogenetic relationships, as well as the genome properties and DDH-based similarities described in Section 3.1, the Octadecabacter clade genomes were divided into 4 groups for subsequent functional analyses and comparisons:

I the polar taxa *O. arcticus* and *O. antarcticus*, as well as the novel polar isolates PAMC 27224 & PAMC 27225

II the temperate taxa *O. temperatus* and *O. ascidiaceicola*

III the phylogenetically more distant *P. jejudonensis* and isolate NH9-P7

IV the temperate isolate E8

Reference clades for the analysis of pan- and core-genome trends were obtained by searching the *Roseobacter* group phylogeny (Figure A.2) for clusters, which branched at a similar distance from the root as the Octadecabacter clade, and contained a similar number (from seven to twelve) of representatives. Four such clades, encompassing a total of 36 representatives, were chosen for comparison, and their members are listed in Table 3.2.

In order to obtain a higher phylogenetic resolution within the Octadecabacter clade, and to test for ambiguities in their relationships, a consensus cluster network representation of 1,513 individual single-copy core-gene phylogenies was constructed. Furthermore, the nine genomes were clustered based on their pairwise Jaccard-distances in gene presence and absence. Both trees are presented in the form of a tanglegram in Figure 3.3.

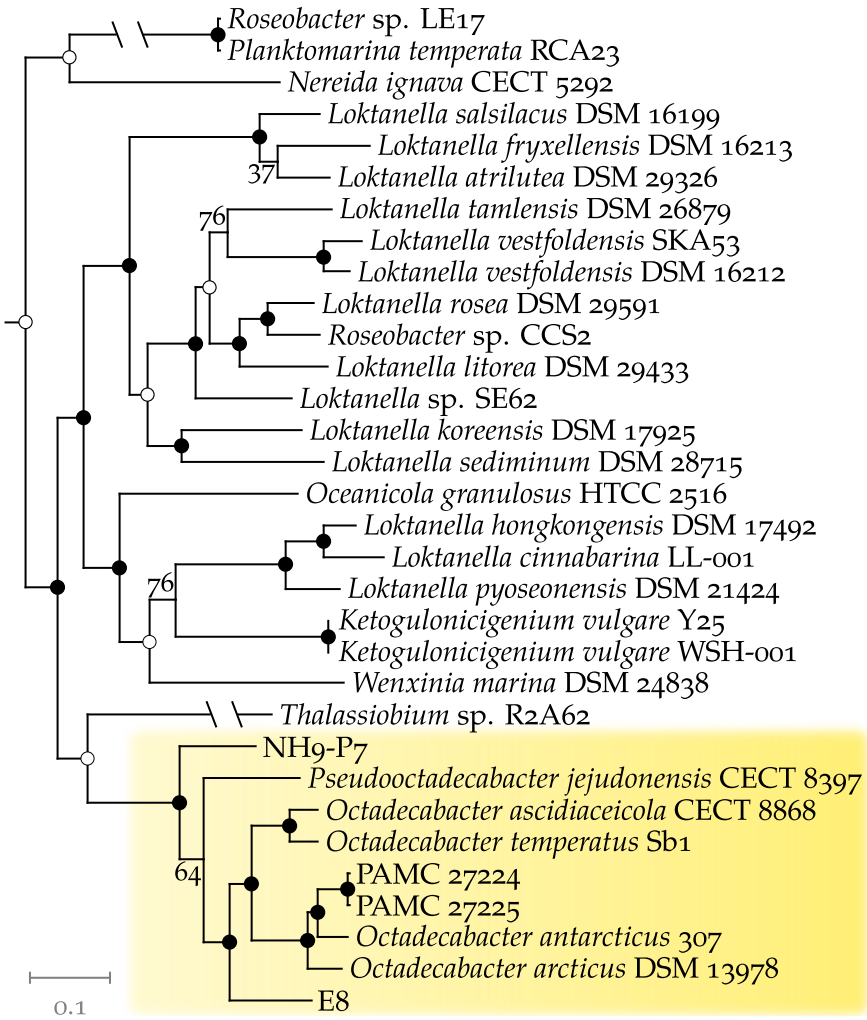


Figure 3.2: MLSA phylogeny encompassing the Octadecabacter clade and its closest relatives. Shown is a subtree pruned out of the larger phylogeny inferred for 115 genomes of the *Roseobacter* group (given in Figure A.2, p. 131). Numbers indicate bootstrap support, with open and closed circles representing values of 90–99 and 100, respectively. Scale bar in expected number of substitutions. Long edges were shortened by 50 %, and are marked with interruptions. The Octadecabacter clade is highlighted.

The network mostly agrees with the MLSA phylogeny in Figure 3.2, but exposes an ambiguous positioning of *O. antarcticus*. A significant amount of single gene phylogenies puts this strain in closer relation to *O. arcticus* than to the PAMC isolates. The internode certainty (IC) and tree certainty (TC) value associated with the respective bipartition are both close to zero, which indicates that both branching patterns occur in a similar number of genes. Relating *O. antarcticus* closer to the PAMC isolates is slightly more frequent than the alternative, as it is associated with a positive IC score. Beside this ambiguity, formation of a polar clade is highly supported (IC score close to 1). *O. ascidiaceicola*, in agreement with DDH and MLSA, is related closest to *O. temperatus* at

Table 3.2: Four monophyletic groups (clades), and their associated *Roseobacter* group genera, which branch at a similar distance from the root as the Octadecabacter clade. Corresponding members are listed for each clade, and were used for comparative pan-, and core-genome analyses (Section 3.3.1).

ASS. GENUS	GENOMES
Loktanella	*L. sediminum* DSM 28715, *L. koreensis* DSM 17925, *Loktanella* sp. SE62, *L. litorea* DSM 29433, *Roseobacter* sp. CCS2, *L. rosea* DSM 29591, *L. vestfoldensis* DSM 16212, *L. vestfoldensis* SKA53, *L. tamlensis* DSM 26879, *L. atrilutea* DSM 29326, *L. fryxellensis* DSM 16213, *L. salsilacus* DSM 16199
Sulfitobacter	*S. guttiformis* KCTC 32187, *S. donghicola* KCTC 12864, *Oceanibulbus indolifex* HEL-45, *S. mediterraneus* KCTC 32188, *S. geojensis* MM-124, *Sulfitobacter* sp. NB-77, *Sulfitobacter* sp. NB-68, *Roseobacter* sp. GAI101, *S. pontiacus* 3SOLIMAR09
Roseovarius	*R. nubinhibens* ISM, *R. atlanticus* R12B, *Roseobacter* sp. AzwK-3b, *Roseovarius* sp. MCTG1562b, *Roseovarius* sp. 217, *R. mucosus* DSM 17069, *Roseovarius* sp. TM1035
Leisingera	*P. gallaeciensis* ANG1, *L. aquimarina* DSM 24565, *L. methylohalidivorans* DSM 14336, *L. caerulea* DSM 24564, *L. aquaemixtae* CECT 8399, *Rhodobacterales* sp. Y4I, *L. daeponensis* DSM 23529, *Roseobacter* sp. SK209-2-6, *Pseudophaeobacter arcticus* DSM 23566, *Roseobacter* sp. MED193

high IC support. Common ancestry of *O. temperatus*, *O. ascidiaceicola*, and the polar clade is less supported than indicated by its bootstrap value in the MLSA phylogeny (Figure 3.2). Likewise, the positioning of isolate E8 directly outside this clade is only slightly more frequent in the individual phylogenies than alternative bifurcations (indicated by a low IC score).

The content-based hierarchical clustering shows a topology similar to the network. Notably, *O. arcticus* and *O. antarcticus* are grouped together, and the PAMC isolates form a separate group inside the polar cluster.

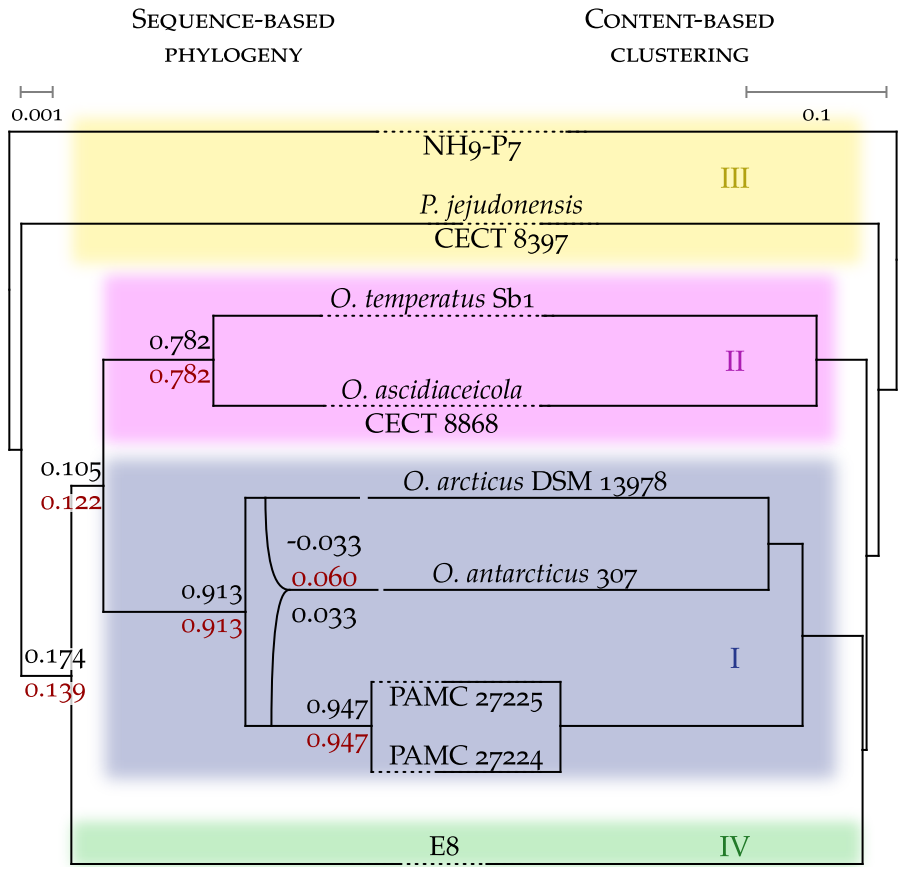


Figure 3.3: Linking core-gene phylogenies to gene content in the Octadecabacter clade. The left is a consensus cluster network representation (⩾20 % frequency) of 1,513 individual Octadecabacter core-gene phylogenies. Note the ambiguous positioning of *O. antarcticus*. Black and red numbers are IC and TC scores, respectively. The right presents a NJ tree computed from the pairwise Jaccard-distances in gene presence and absence. See Section 2.7 for methods. Scales in expected number of substitutions (left) and distance (right). Both trees were manually rooted, using isolate NH9-P7 as outgroup. Coloured backgrounds and roman numerals indicate groups for gene content comparison (see main text).

3.3 GENERAL GENE CONTENT

3.3.1 *Pan-genome analysis*

Pan- and core-genome sizes by number of included genomes are shown in Figure 3.4 for the Octadecabacter clade. The fitted power law curve, with an exponent of 0.512, does not reach saturation. Their extrapolated core-genome size is 1,496. Table 3.3 provides standard deviations (SDs) for both parameters, as well as fitting results for lineages within other *Roseobacter* group genera, which branch at a similar phylogenetic depth. Over all roseobacter genomes used in this study, the pan-genome exponent is 0.66, and the estimated core-genome size is 287 (Figure 3.5, see bottom of Table 3.3 for SDs). Concerning the latter analysis, it should be

noted that the sample size of 600 at each number of genomes is still representative of the underlying distribution, although small compared to the number of possible combinations (e. g. $\binom{114}{60} \approx 13 \cdot 10^{33}$) [188].

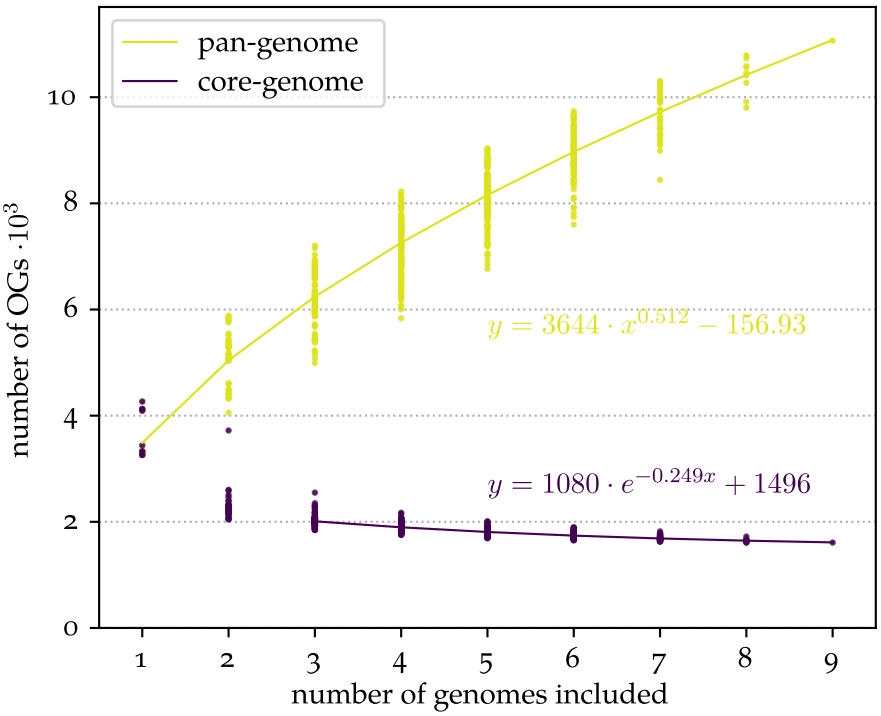


Figure 3.4: Number of pan- and core-OGs for all possible permutations of the indicated number of genomes from the Octadecabacter clade. Curves were fitted based on the median values.

Table 3.3: Exponent b of the pan-genome, and extrapolated core-genome size of the core-genome curve fits are provided for the Octadecabacters, as well as lineages within other *Roseobacter* group genera (see Table 3.2). Values represent estimate ± SD. n: number of genomes

ASS. GENUS	n	b	CORE
Octadecabacter	9	0.51 ± 0.03	1,496 ± 17
Loktanella	12	0.644 ± 0.025	1,162 ± 20
Sulfitobacter	9	0.632 ± 0.018	1,477 ± 24
Leisingera	9	0.526 ± 0.024	2,027 ± 34
Roseovarius	7	0.525 ± 0.025	1,357 ± 369
all genomes	114[a]	0.660 ± 0.001	287 ± 6

[a] as in Table 2.2, excluding *Parvularcula bermudensis* (outgroup genome), and *Phaeobacter* sp. LSS9

Koonin and Wolf [189] propose to divide the pan-genome into three categories, based on its frequency in the constituent genomes:

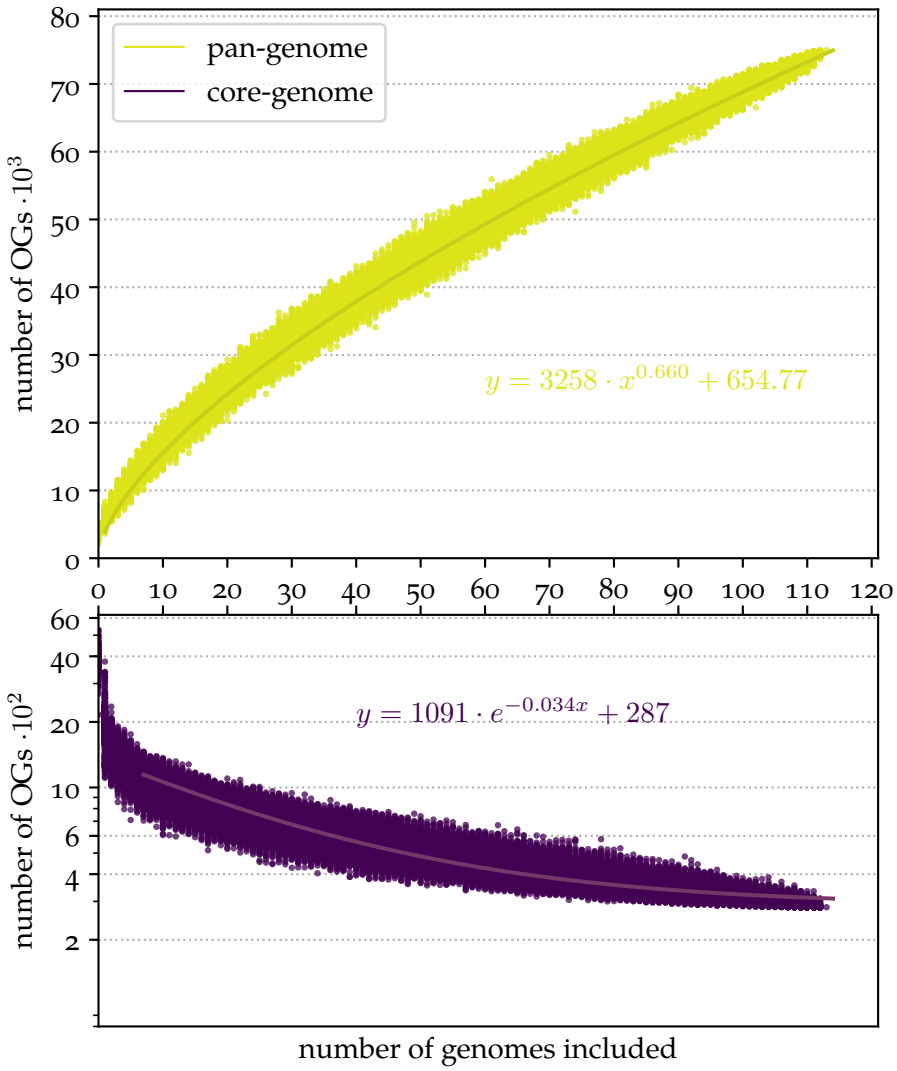


Figure 3.5: Sizes of pan- (top) and core-genome (bottom) for at least 600 random combinations of the indicated number of genomes from the *Roseobacter* group (Table 2.2). Curves were fitted based on the median values.

i the *core-genome*, consisting of genes present in all or the vast majority of comparison organisms,

ii the *cloud-genome*, containing genes which are less frequent than the core-genome, but occur in multiple of the comparison genomes, and

iii the *shell-genome*, i. e. genes exclusive to one or a few organisms.

In the Octadecabacter clade, with nine sequenced representatives, sensible boundaries for these categories are $n = 9$ for the core-, $1 < n < 9$ for the cloud-, and $n = 1$ for the shell-genome, with n being the number of genomes sharing a respective OG. Absolute CDS counts of each category are provided in Figure 3.6 for each compared genome individually. In all cases, the core-genome, at 1,609 OGs, makes up the

largest respective fraction of CDSs. The polar Octadecabacter genomes are larger and contain, on average, ≈800 more CDSs than the temperate strains (apparent also from Table 3.1). Most of these CDSs fall into the shell-genome, i. e. are exclusive to a single representative (note that this observation is partially masked in the PAMC-isolates by their high pairwise similarity, i. e. close relation, to each other). *O. arcticus* possesses the largest absolute shell-genome among the currently sequenced representatives (making up ≈30 % of its CDSs).

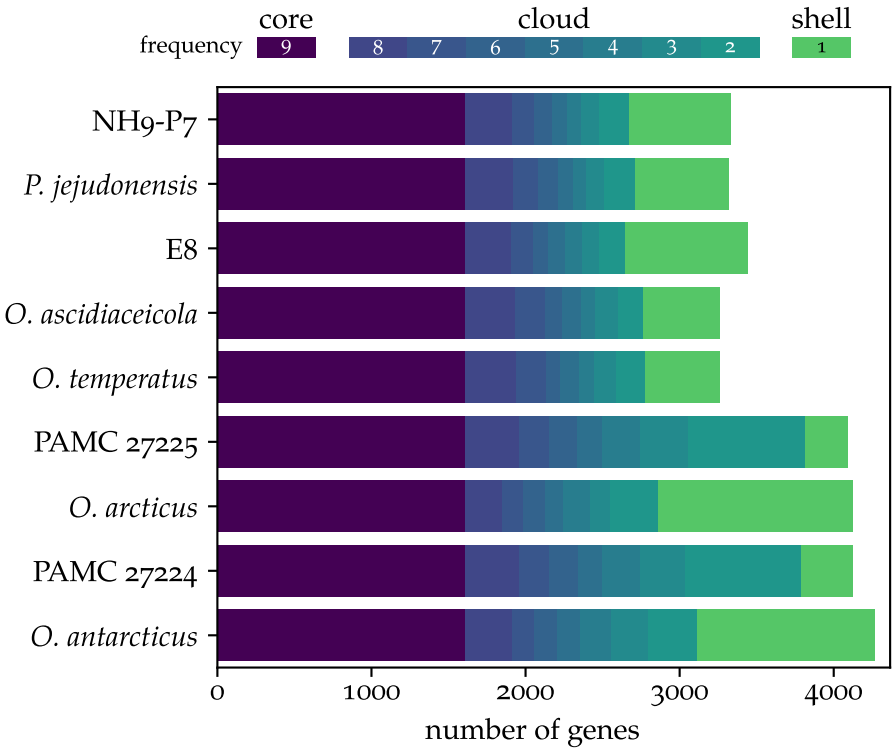


Figure 3.6: Genome category sizes (in number of unique genes, i. e. discounting paralogues) for each Octadecabacter clade genome. Category boundaries in number of genomes n were $n = 9$ for core-, $1 < n < 9$ for cloud-, and $n = 1$ for shell-genome.

3.3.2 *Distribution of COG categories*

The number and frequency of genes in different COG categories were analysed for the four Octadecabacter clade groups defined in Section 3.2. The fact that polar Octadecabacter clade members (group I) possess larger genomes than the others (Table 3.1) has to be taken into account when performing a direct comparison. Generally, three distinct trends of COG category size change are possible by comparing larger with smaller genomes:

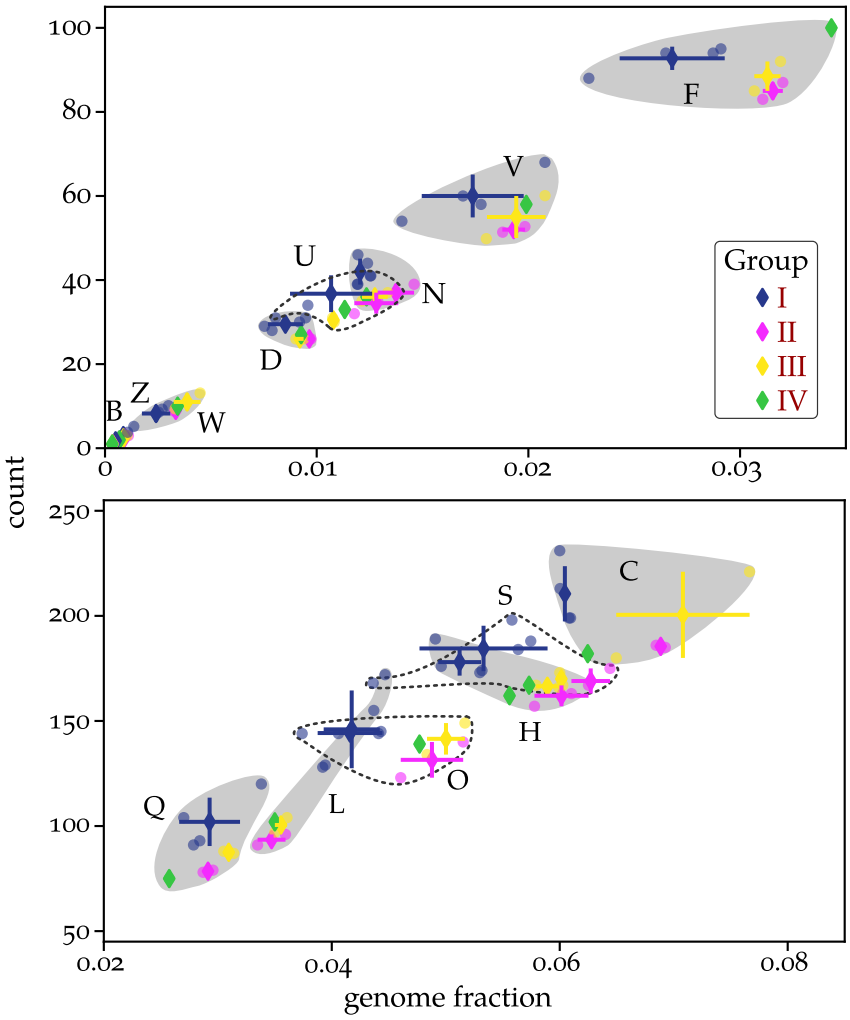
i the number of genes in a category may remain the same, or decrease, which concomitantly decreases this category's fraction of

the total COG content (which will be referred to as a negative size correlation from here on),

ii the number of genes in a category may increase linearly with the total number of genes, and this category will therefore occupy the same genome fraction in small and large genomes, and

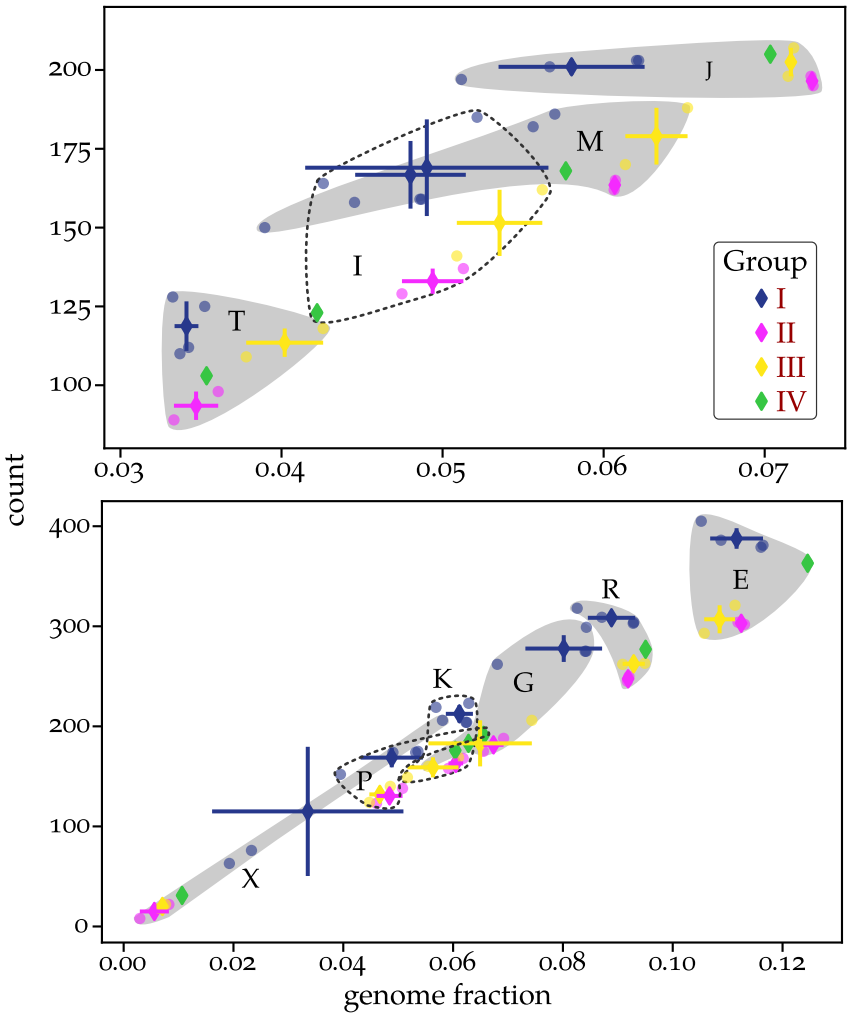
iii the number of genes in a category may increase super-linearly with the total number of genes, leading to this category occupying a larger fraction of the total COG content.

Both, absolute numbers and category fractions are provided for the four groups in Figure 3.7. Visually, each COG category was assigned to one of the three behaviours described above, summarised in Table 3.4.



(a) COG categories B, C, D, F, H, L, N, O, Q, S, U, V, W, and Z

Figure 3.7: Mean absolute numbers of COGs in each category are plotted against its fraction of total COG content as diamonds. Dots show individual values of the constituent genomes. Groups are as defined in Section 3.2. Bars indicate SD).



(b) COG categories E, G, I, J, K, M, P, R, T, and X

Figure 3.7: COG category distributions in the Octadecabacters (continued)

Table 3.4: COG category trends in polar Octadecabacter clade genomes compared to the temperate strains. Trends were derived from Figure 3.7. For their definitions, see the main text.

negative correlation with genome size	
C	Energy production and conversion
F	Nucleotide transport and metabolism
H	Co-enzyme transport and metabolism
J	Translation, ribosomal structure, and biogenesis
M	Cell wall, membrane, and envelope biogenesis
O	Posttranslational modification, protein turnover, and chaperones

Table 3.4: continued

S	Function unknown
	linear increase with genome size
E	Amino acid transport and metabolism
I	Lipid transport and metabolism
K	Transcription
N	Motility
P	Inorganic ion transport and metabolism
Q	Secondary metabolite biosynthesis, transport and metabolism
R	General function prediction only
T	Signal transduction mechanisms
U	Intracellular trafficking, secretion, vesicular transport
V	Defence mechanisms
	super-linear increase with genome size
G	Carbohydrate transport and metabolism
L	Replication, recombination, and repair
X	Mobilome (prophages, transposons)
	unclear trend
B	Chromatin structure and dynamics
D	Cell cycle control, cell division, chromosome partitioning
W	Extra-cellular structures
Z	Cytoskeleton

3.4 GROUP-SPECIFIC GENES

specific – present in all members of one group and absent in all other groups

Figure 3.8 gives an overview of the number of genes specific to each group defined in Section 3.2, and any of their combinations. `po2group_stats.pl` estimates a core-genome size of 1,609 for the nine genomes. At 90, group I (polar Octadecabacters) has the highest number of specific genes (discounting the 795 accessory genes of isolate E8). Paralogous genes in the reference genome, listed by `po2group_stats.pl` as multiple entries, were only counted once. The individual genes specific to each group are provided in Tables A.2 to A.5, Appendix A.4, p. 135.

Fifty-five of the 90 genes specific to polar Octadecabacters are co-localised at 10 positions, with cluster sizes ranging from 2 to 14. Sev-

enteen genes are annotated as hypothetical proteins (or their functions are not clearly specified), and 5 as transposases or recombinases.

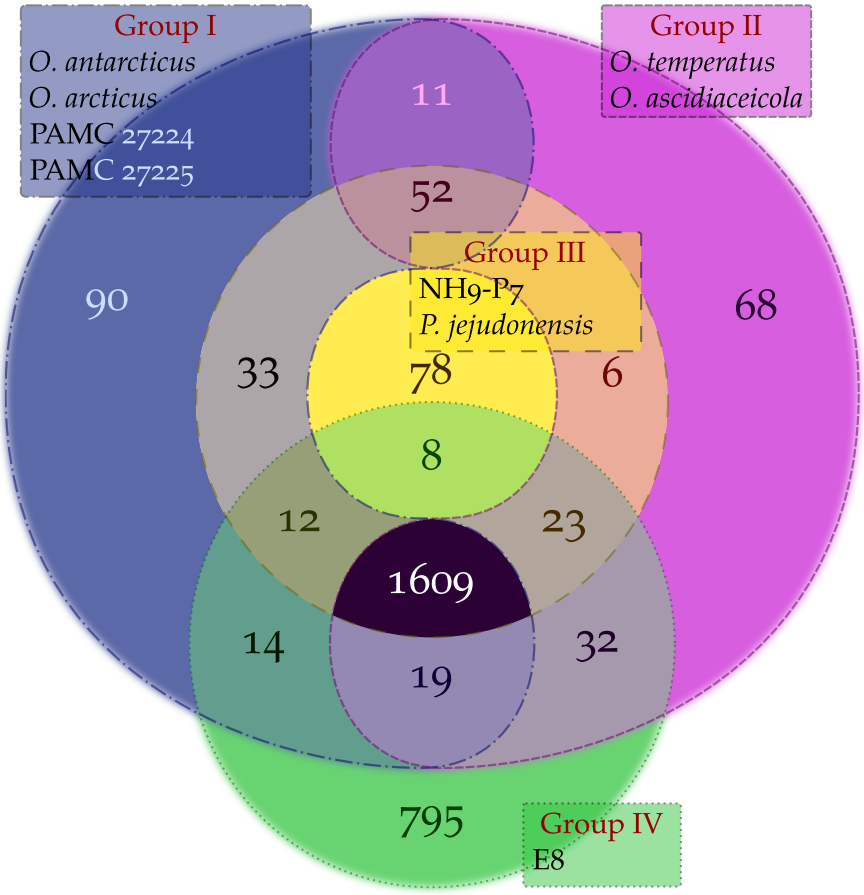


Figure 3.8: 4-set Venn diagram showing number of genes shared between and specific to each group defined in Section 2.10.2.

Cases in which multiple genes are encoded directly adjacent to each other, in conserved gene clusters, are of particular interest, as adjacency often entails functional interaction or interrelation [190]. Therefore, these cases will be discussed concomitantly in the following sections.

Since the adaptation of polar Octadecabacters (group I) to their extreme habitat is of main interest, focus will be placed on the description and interpretation of genes specific to this group.

Comparison group IV consists of isolate E8 as the only member, and its 795 specific genes therefore constitute the accessory genome of this strain. As a generalisation from one individual is not meaningful, discussion of these genes will be omitted here.

3.4.1 *Genes specific to polar Octadecabacters*

As stated above, 55 of the 90 genes specific to polar Octadecabacters are co-localised at 10 positions. The corresponding gene products include

several transcriptional regulators such as MarR, stress-related proteins, a xanthorhodopsin, dehydrogenases, and enzymes for the degradation of various substrates (Figures 3.9 to 3.18). Seventeen genes are either annotated as hypothetical proteins, or their functions are not clearly specified.

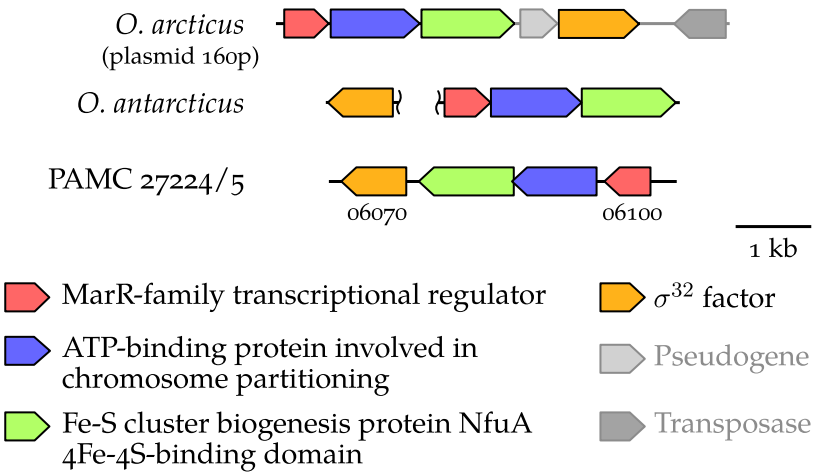


Figure 3.9: MarR-associated gene cluster in polar Octadecabacters

Polar Octadecabacter-specific regulatory proteins include a member of the MarR-family (Figure 3.9). MarR trancriptional regulators are associated, among other factors, with oxidative stress in *E. coli* [191]. The corresponding gene is usually found adjacent and divergently transcribed to the target genes it regulates, and through binding the intergenic region, it represses both its own and its targets' transcriptions [191]. The current model of its inactivation through oxidative stress is as follows: oxidation of one or multiple MarR cysteine residues leads to conformational changes, which impair DNA-binding of this repressor and subsequently lead to transcription (activation) of its target operon [191]. The corresponding oxidative stress signal was recently found to be brought about by copper ions, released upon cell envelope stress, in *E. coli* [192]. MarR may also be involved in metabolic regulation, where it is activated by binding to a specific substrate, and consequently acts as transcriptional activator of the respective catabolic pathway, binding upstream of the respective genes [193].

It has not yet been experimentally tested whether the conserved, polar Octadecabacter-specific cluster formed by MarR and its three neighbouring genes is actually under MarR regulation. However, the corresponding gene functions, which are involved in stress response indicate this: in cases in which it is induced by stress, MarR is often associated with efflux pumps [191], and the downstream neighbouring gene octa_06090 has >70 % identity to a sodium:proton antiporter in a reference member of the *Rhodobacteraceae*. One of the factors severely affected by oxidative stress is sulphur biochemistry [49], and fittingly, the next downstream neighbour in polar Octadecabacter MarR-gene clus-

ters encodes an NfuA-like gene product. NfuA assists FeS-cluster biogenesis, likely by binding newly synthesised FeS-clusters and shuttling them to their targets [194]. It can thus shield these FeS-clusters from the intracellular environment, thereby protecting them from oxidative stress [194]. For example, aconitase B of the tricarboxylic acid (TCA) cycle is one important NufA target in *E. coli* [194].

The final gene found in conjunction with this MarR-associated gene cluster is a sigma factor related to σ^{32}, although it is not always in direct neighbourhood. This sigma factor is involved in the heat-shock response of *E. coli*, but also regulates transcription of additional genes [195]. For polar Octadecabacters, it may exert a protective function by inducing expression of chaperones. Notably, it may also be involved in maintenance of proper sulphur biochemistry through induction of thioredoxin expression [195].

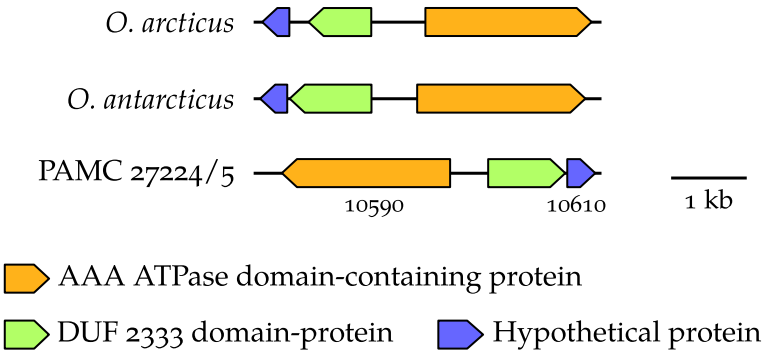


Figure 3.10: AAA ATPase specific to polar Octadecabacters

A potentially stress response-associated gene specific to polar Octadecabacters is an AAA ATPase (Figure 3.10). These are found in all domains of life and generally induce conformational changes in proteins upon adenosine triphosphate (ATP)-phosphorylation [196]. They act in a wide range of cellular processes, most notably in protein unfolding at the proteasome lid, dis-assembly of protein aggregates and complexes, and membrane protein extraction for subsequent degradation [196]. This unfolding property may be particularly important for polar Octadecabacters, due to the low resident temperatures of their habitats. As described in Section 1.2, the cell membrane becomes more rigid at low temperature, which may necessitate this additional factor for proper membrane protein extraction and degradation. Although protein stability is generally decreased as adaptation to low temperatures, the specific properties of some proteins might still require accessory functions for their successful degradation, such as the AAA ATPase.

Another, apparently stress-associated, component found only in polar Octadecabacter genomes affects cellular selenium biochemistry (Figure 3.11). Selenium is an important trace element, which may be incorporated into proteins (mostly in the form of selenocysteine) and nu-

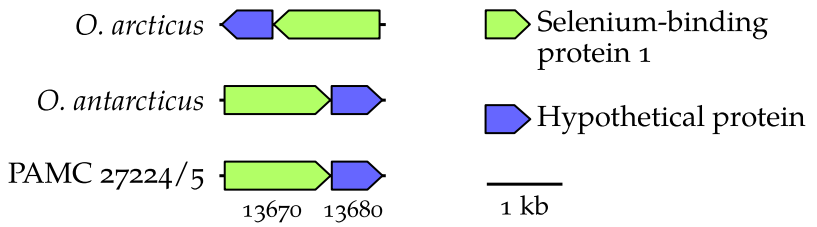


Figure 3.11: Selenium-binding protein specific to polar Octadecabacters

cleotides [197]. All Octadecabacter clade members seem to synthesise selenophosphate (via the product of *selD*) and to use 2-selenouridine-modified tRNA (*selU* gene). Free selenium species induce oxidative stress, most importantly by binding glutathione molecules, thereby forming O_2^- radicals [198]. Detoxification happens via reduction by thioredoxin [197], or, at lower rate, spontaneously-formed elemental selenium (Se^0). Selenium-binding proteins (SBPs) may bind free selenium species through exposed thiol groups and shuttle them to selenophosphate synthase, thereby shielding them from the cellular environment [197]. However, other thiol-containing cellular enzymes, such as 3-mercaptopyruvate sulfurtransferase (3-MST) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), bind and release free selenium effectively, and could therefore suffice as shuttles to supply selenophosphate synthase with selenium [197, 199]. The function of SBPs consequently remains elusive to this day [197].

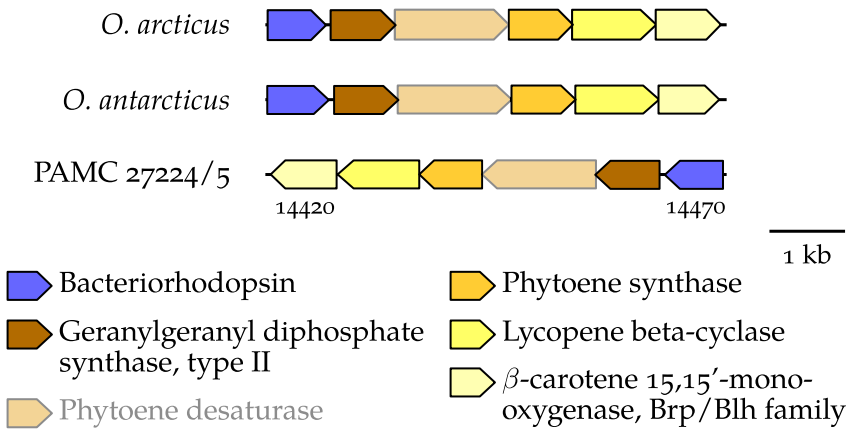


Figure 3.12: Xanthorhodopsin gene cluster specific to polar Octadecabacters

All polar Octadecabacter strains contain a rhodopsin-coding gene cluster (Figure 3.12), enabling potential photoheterotrophic life strategies. Bacterial rhodopsins were originally discovered in a marine fosmid clone [200], and it is now clear that they are more abundant than photochemical reaction centres in most marine habitats [201]. Vollmers et al. described a new subgroup of rhodopsins, the so-called group II xanthorhodopsins in *O. arcticus* and *O. antarcticus*, which they associated with cold and saline habitats [77]. The gene locus organisation of

this group of rhodopsins is also conserved, with the five genes needed for synthesis of the rhodopsin's chromophor retinal localised downstream of the opsin gene [77] Phylogenetic reconstruction of the PAMC isolate-derived rhodopsins (see Figure A.6 on page 168), as well as the organisation of the gene locus (Figure 3.12) revealed that these belong to the same group. The fact that in the present analysis, group II xanthorhodopsins were found to be specific for all polar Octadecabacters, supports the association by Vollmers et al. While this type of phototrophy is relatively rare among *Roseobacter* group members (eight out of 115 genomes used presently, including one proteorhodopsin), xanthorhodopsins are generally exclusive to polar roseobacters, and occur in the majority of them, as visualised in Figure A.5 (p. 167).

The cellular role of xanthorhodopsins is currently still unclear, and may differ greatly between species [202]. Generally, xanthorhodopsins, like their more abundant sister-group proteorhodopsins, utilise light energy to translocate protons across the cell membrane, thereby generating a proton-motive force [200, 203]. This force could be used for ATP-synthesis, ion transport, or flagellar propulsion [202]. Importantly, in *E. coli*, proteorhodopsin was not able to create a greater membrane potential than the respiratory chain [204], consequently not adding to the proton-motive force and the observed growth speed. However, it was able to recover the membrane potential when the respiratory chain was blocked, implying a protective function for the cell during respiratory stress [204].

Experiments on *Vibrio* species implied involvement of proteorhodopsin in starvation survival [205], and Vollmers et al. hypothesised a similar function for the xanthorhodopsins of the sea ice residents *O. arcticus* and *O. antarcticus* [77].

The largest gene cluster specific for polar Octadecabacters was found to consist of 15 syntenic genes in the PAMC isolates, but two separate regions, respectively, in *O. arcticus* and *O. antarcticus* (Figure 3.13). In this cluster, a LacI-family transcriptional regulator is found adjacent and divergent to multiple partly overlapping genes, indicating that the cluster constitutes an operon. This family usually regulates carbohydrate metabolism [206], which concurs with the downstream presence of a CUT 1 family carbohydrate ABC-transporter. The CUT 1 family may transport a broad range of substrates, such as polyols, oligosaccharides, or glycerol-phosphate [207]. An adjacent dihydroxyacetone (DHA) kinase is involved in glycerol degradation [208]: DHA is formed from glycerol through the action of a dehydrogenase, and subsequently phosphorylated to form DHA phosphate (DHAP), a glycolysis intermediate. Notably, the Octadecabacters do not possess the respective dehydrogenase (instead, all Octadecabacters degrade glycerol via phosphorylation using a glycerol kinase, and subsequent oxidation to DHAP). Instead, DHA may be imported from the environment (e. g. through the ABC-transporter in question), as has been observed for other hyper-

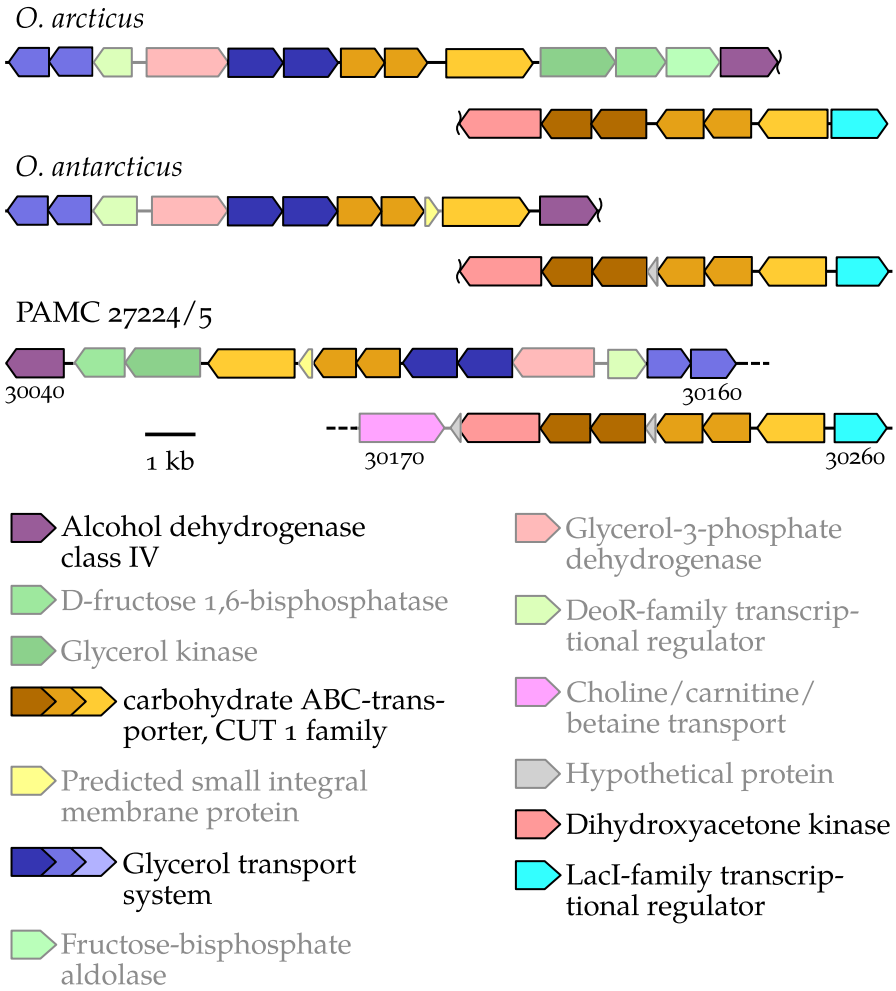


Figure 3.13: Cluster of genes associated with uptake and degradation of carbohydrates, specific to polar Octadecabacters

saline communities: in short, primary producers synthesise glycerol as osmoprotectant, which heterotrophs then take up for degradation. A part of the imported glycerol may be secreted again, after oxidation, as DHA [209]. DHA thus constitutes an overflow product [209]. The situation may be similar in the sea ice habitats of polar Octadecabacters, since they live in close association with primary producers, and salinity often reaches near-saturation levels in brine channels (Section 1.2).

The remaining genes in this region possess similar functions: subunits of a glycerol ABC-transporter are localised beside an incomplete CUT 1 system, and both are specific to polar Octadecabacters. Furthermore, with the presence of a glycerol kinase and a glycerol-3-phosphate dehydrogenase, the complete pathway of glycerol degradation is found in this locus, under the control of a DeoR-family transcriptional regulator. The latter three components are not exclusive to polar isolates, as orthologues of these can be found in all Octadecabacter clade genomes sequenced to date.

In conclusion, this locus seems to be involved in glycerol degradation in polar Octadecabacters. This degradation may also be regulated in response to osmotic and temperature changes, since glycerol is synthesised as both osmo- and cryoprotectant by many algae.

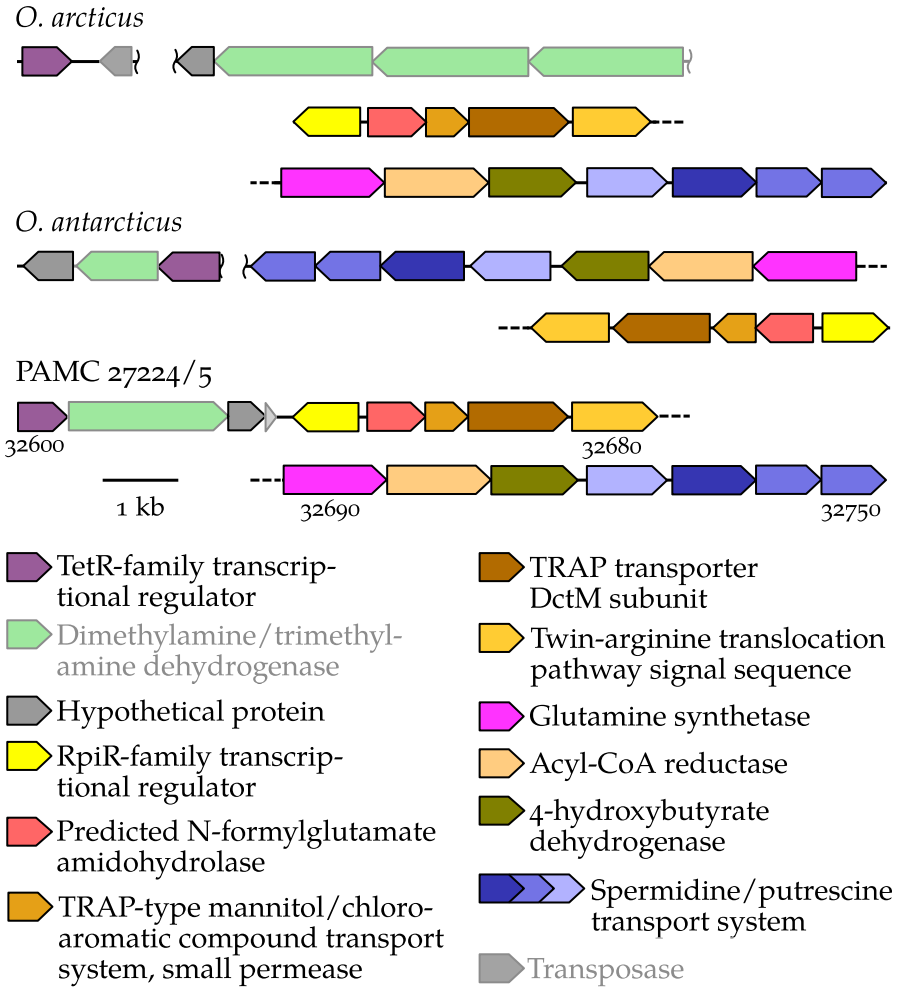


Figure 3.14: RpiR- and TetR-associated gene clusters specific to polar Octadecabacters

Another large cluster specific for polar Octadecabacters was found to consist of a set of eleven syntenic genes, preceded by an inversely transcribed RpiR-family regulator (Figure 3.14).

The role of the transcriptional regulator has not been characterised in this group. RpiR was first described to repress ribose phosphate isomerase B of the pentose phosphate pathway in *E. coli* [210], whereas it serves as an activator in *Staphylococcus aureus* [211]. Generally, the family is involved in regulation of sugar catabolism [211, 212]. This could functionally correlate with a downstream tripartite ATP-independent periplasmic (TRAP) transport system encoded in the same gene cluster. The twin-arginine translocation pathway signal sequence is annotated as the *dctP* component of a TRAP system in *O. arcticus* and *O. antarcti-*

cus, thus constituting a complete system. These transport a wide range of substrates in different organisms, including ectoin and carboxylic acids [213]. It is important to note that substrate specificity can only reliably be determined in vitro [214]. However, uptake of mannitol, as suggested by automatic annotation, or a similar compound would provide intermediates for use in either glycolysis, or the pentose phosphate pathway.

Apart from the TRAP transporter, genes for a spermidine/putrescine ABC-transport system are also located at this site. All Octadecabacters possess several copies of this transporter, and are able to degrade putrescine via the putrescine utilisation pathway (*puu* operon [215]), but its corresponding biosynthetic pathway is absent. Polyamines generally provide protection against a number of stress factors [216], but also serve as a nitrogen source. Indicative of a primary role as nitrogen donor in this context is an upstream glutamine synthetase (GS) gene within the same cluster, the product of which constitutes the primary mechanism of nitrogen assimilation [217].

A predicted N-formylglutamate amidohydrolase within the same cluster could catalyse the last step in histidine degradation (see Section 3.5.4 for more detailed information), which yields glutamate and formate. The former serves as substrate for GS, thereby increasing the capacity for nitrogen uptake.

In addition, a TetR-family regulator can be found upstream, in relatively close proximity to this gene cluster in the PAMC isolates, but at varying locations and genomic contexts in the polar *Octadecabacter* type strains. This family of regulators can affect a wide range of functions, including metabolic homoeostasis [218]. It is, however, most often associated with the regulation of efflux pumps under conditions of cellular stress [218].

In conclusion, the genes in this cluster may be primarily involved in the regulation of purine and pyrimidine synthesis, since both glutamine, and intermediates of the pentose phosphate pathway are required as substrates in these pathways.

Three of the polar Octadecabacter-specific genes were found to lie in close proximity, albeit not directly adjacent to each other, in most of the four genomes (Figure 3.15). Among them is a transporter of the BASS family, which imports (primarily) bile acids in conjunction with sodium ions, and has been well characterised in animals [219]. The family is also known to transport other substrates, such as steroids and their derivatives [219]. Structural data and general characterisations of this family are sparse in prokaryotes [220], and we are consequently lacking detailed information regarding their functions and substrates in this group of organisms. Generally, however, bile salts frequently occur in the environment, and can be used as the sole carbon source by some bacteria, although the genetics and regulation behind this process are currently poorly characterised [221]. Similarly, *myo*-inositol widely oc-

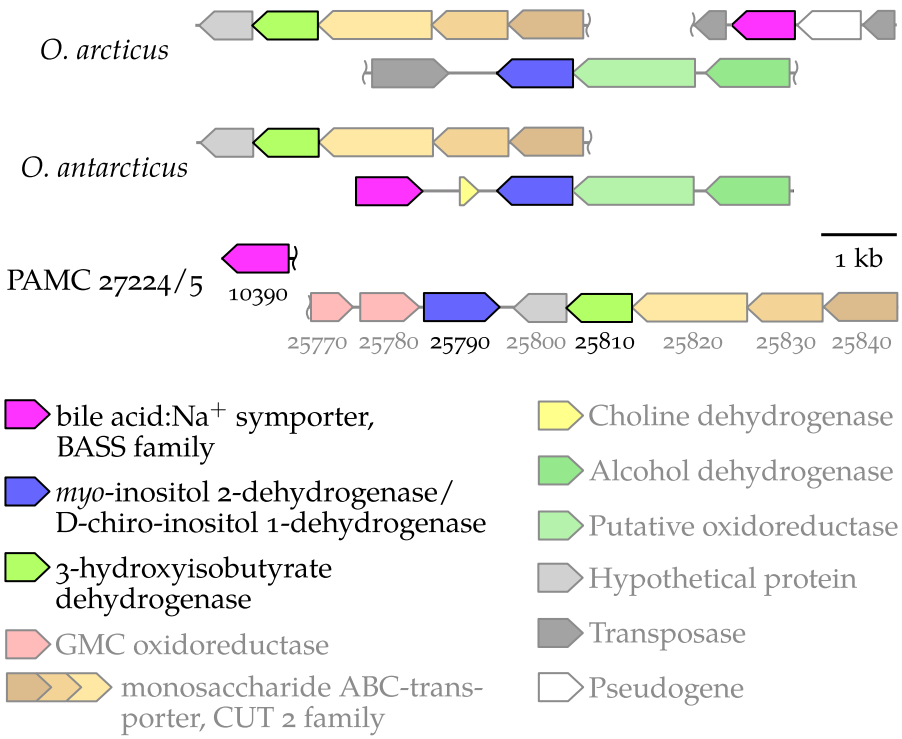


Figure 3.15: Genes specific to polar Octadecabacters, which may be involved in degradation of diverse substrates

curs in natural habitats, and is readily degraded by diverse bacteria to DHAP and acetyl-coenzyme A (CoA) (Ac-CoA) [222–225]. *myo*-inositol dehydrogenase, which was found to be specific for polar Octadecabacters, catalyses the first step in *myo*-inositol degradation [226]. Usually, the genes needed for its catabolism are organised in the *iol*-operon [226], and an orthologous form of this operon is present in all currently sequenced Octadecabacters except for isolates E8 and NH9-P7. The polar-specific dehydrogenase is non-paralogous to its homologue in this operon, and was therefore likely gained by a common ancestor of the polar Octadecabacters through lateral transfer.

Finally, a polar Octadecabacter-specific 3-hydroxyisobutyrate dehydrogenase gene is positioned downstream and adjacent to a (non-specific) carbohydrate ABC-transporter. Its gene product is involved in valine degradation, which ultimately results in succinyl-CoA. CUT 2-family transporters are usually specific for monosaccharides [207], and the transport system found at this location therefore likely is not involved in either valine, or *myo*-inositol uptake.

Without appropriate experimental examination, a detailed function of these three specific genes remains unknown. However, they generally seem to be involved in the uptake and degradation of carbohydrates, thereby broadening the spectrum of substrates polar Octadecabacters can utilise for their growth.

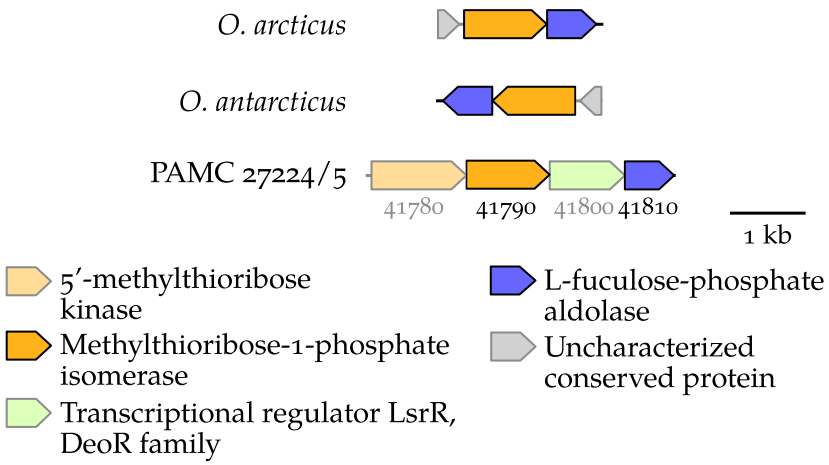


Figure 3.16: *mtnA* and *fucA* in polar Octadecabacters

An *mtnA* gene (and its immediate genomic context) was also found to be specific for polar Octadecabacter isolates (Figure 3.16). Its product, methylthioribose-1-phosphate isomerase, participates in the methionine salvage pathway, which may not be complete in these organisms (see Section 3.5.3).

In *O. arcticus* and *O. antarcticus*, this gene is directly adjacent to the group-specific *fucA*, coding for L-fuculose-phosphate aldolase. This enzyme is involved in L-fucose and D-arabinose degradation, and specifically converts L-fuculose-1-phosphate to DHAP [227]. The two genes could therefore broaden the range of substrates polar Octadecabacters can use for growth, thereby reflecting diversity as well as scarcity of nutrients in their natural habitats.

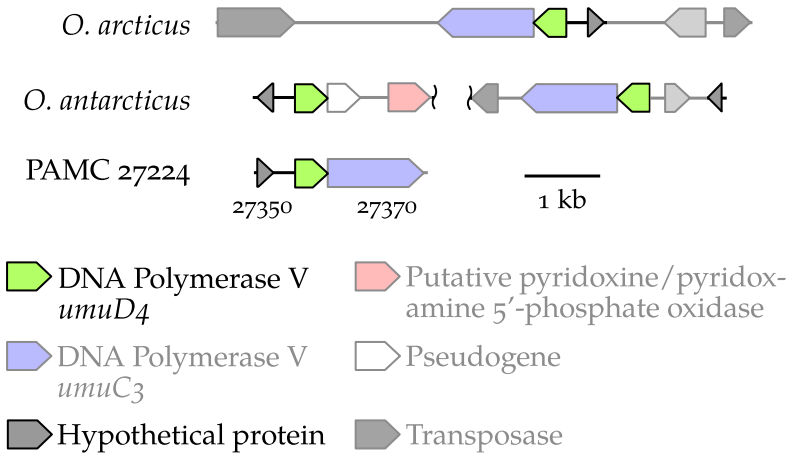


Figure 3.17: Genes encoding DNA polymerase V specific to polar Octadecabacters

Among the OGs specific to polar Octadecabacters are genes coding for a DNA polymerase (Figure 3.17). DNA polymerase V belongs to the Y-family of polymerases, and is involved in error-prone repair

of DNA damage, most notably translesion DNA synthesis (TLS) [228]. A heterotrimeric $UmuD'_2C$ complex forms the functional polymerase. While no annotated *umuC* gene was found in the PAMC isolate 27225, and this gene can therefore currently not be considered group-specific, this is likely due to the high fragmentation of that genome (Table 3.1).

Generally, DNA Pol V is strongly induced after ultraviolet (UV) irradiation, which is seasonally intense in the polar regions [50], and exacerbated by decreasing atmospheric ozone content [229]. Induction happens at a late stage of the stress response, and this is interpreted as a compromise between its ability to perform TLS and its low fidelity (it possesses no 3'-5' exonuclease activity) [228]. Importantly, the high mutability was also shown to constitute an adaptive advantage: when Yeiser et al. co-incubated *umu*-deficient and wildtype *E. coli*, the wildtype always outgrew and ultimately eliminated the mutant [230]. The multiple copies of *umuC/D* in the polar Octadecabacter genomes may therefore constitute less the result, and more a means of adaptation to a new environment.

The short hypothetical protein found in conjunction with both *umu* genes contains a domain of unknown function (DUF) 1127, and has orthologues in multiple other *Roseobacter* group bacteria.

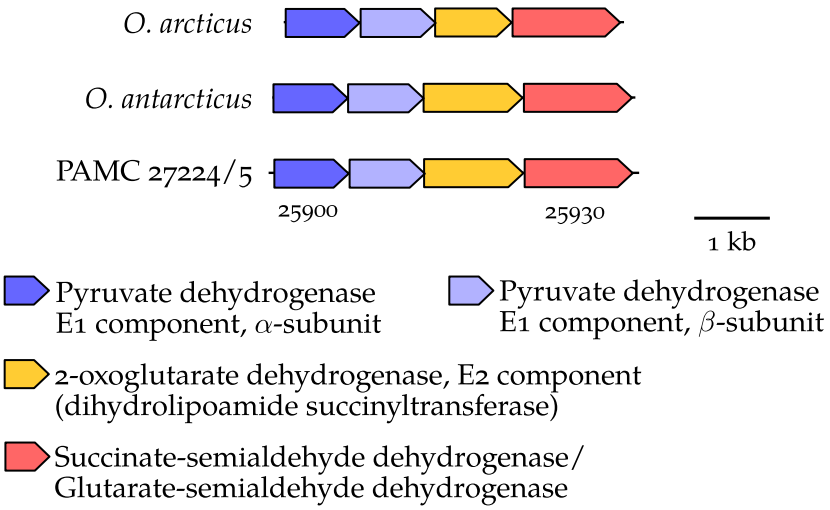


Figure 3.18: Dehydrogenase gene cluster specific to polar Octadecabacters

Four dehydrogenase genes which are involved in the TCA cycle, the central component of cellular metabolism, are specific for polar Octadecabacters (Figure 3.18). These seem to represent additional variations of common TCA cycle components, which are non-orthologous to those of the other Octadecabacters. Although redundant, the fact that the four genes are group-specific and conserved in order suggests that they are part of a specific regulatory stage in polar Octadecabacters.

According to automatic annotation of isolate PAMC 27224, two genes code for the E1 component of pyruvate dehydrogenase, although these are usually difficult to distinguish from the respective subunits of ace-

toin dehydrogenase. Both enzyme complexes form Ac-CoA, the former from pyruvate, thereby linking glycolysis with the TCA cycle, and the latter from acetoin [231].

2-Oxoglutarate (=α-ketoglutarate (α-KG)) dehydrogenase forms succinyl-CoA from α-KG.

The three enzyme complexes are structurally related. Each consists of three components (named E1-3), the last of which is freely shared between all [231]. However, none of the other components were found as orthologues specific for the polar Octadecabacters.

Succinate-semialdehyde dehydrogenase produces succinate from its semialdehyde, which in turn derives from, e. g. putrescine degradation [215].

The TCA cycle is a central component of cellular metabolism, both providing reducing equivalents for oxidative phosphorylation, as well as contributing important precursor molecules to biosynthesis, among other functions [232]. Flux models of *E. coli* predict that this central pathway, along with the amino acid and purine synthesis pathways and a few others, dominates metabolic flux in the cell at most times, a network which is termed the high-flux backbone (HFB) [233]. It is this HFB through which the cell reacts to changes in nutrient availability, by changing the flux through individual reactions, and (de-)activating relevant auxiliary pathways [233]. Consequently, concomitant expression of the four dehydrogenase genes, induced under specific circumstances, may serve metabolic steering in the polar group, and may be part of a cellular state or strategy, which is not encountered in the other Octadecabacters. For example, pyruvate/acetoin dehydrogenase would increase flux through the TCA cycle by providing Ac-CoA, α-KG dehydrogenase would steer flux towards succinate (away from glutamate synthesis), which would also be provided by succinate-semialdehyde dehydrogenase. This configuration would consequently either lead to increased generation of reducing equivalents and ATP through the TCA cycle, or feed into biosynthesis pathways, which consume succinyl-CoA, most importantly cobalamin and other tetrapyrroles [234].

The polar-specific genes, which are not part of consecutive regions, generally mirror the adaptations described above, i. e. they are involved in the cell's response to oxidative stress, broaden the range of usable substrates, or convey similar functions. For example, in addition to NfuA, two further group-specific enzymes are involved in sulphur biochemistry: one protein of the cysteine desulphurase-family (octa_09620), and a thiol-disulphide oxidoreductase of the DCC family (octa_14550). The former transfers sulphur from cysteine to a range of possible targets, and may be involved in biosynthetic processes of thiamine, biotin, FeS-clusters, thionucleosides, and many other compounds [235]. The latter class of enzymes manages the redox balance of thiol-disulphides of the intra- and extracellular spaces [236]. In their description of the DCC family, Ginalski et al. noted this family's non-

universal taxonomic distribution, and high evolutionary distance from the related thioredoxin-like families, which they took as an indicator of a specific, however still undetermined, function [237].

Besides coping with an increased presence of reactive oxygen species, bacteria may adapt to alleviate the effects of higher oxygen solubility at low temperatures by consuming more of it in enzymatic reactions [39]. Beside possessing, on average, slightly more genes coding for (oxygen consuming) dioxygenases than the temperate clade members, polar Octadecabacters also possess group-specific enzymes, which convey this function. One of these is the taurine dioxygenase TauD (octa_12020), which cleaves taurine to sulphite and aminoacetaldehyde under consumption of molecular oxygen [238]. They further possess a group-specific glycine/D-amino acid oxidase (octa_06740), which partakes in the degradation of amino acids through oxidative deamination [239].

As mentioned earlier, the polar Octadecabacters may frequently encounter situations of osmotic stress. Fittingly, they possess a group-specific cation/H^+ antiporter of the CPA1 family (octa_12300), which is known to confer increased salt tolerance through increased export of cations from the cell [240]. This type of anitporter was also shown to be present in the genomes of other cold-adapted bacteria [39].

Regarding substrate utilisation, multiple individual components of uptake systems for, e. g. phosphate, ribose, or simple sugars were found specific for the polar group, in addition to those contained in the loci described above. Furthermore, a group-specific 4-α-glucanotransferase (octa_10680) may avail them maltose [241], or storage polysaccharides such as glycogen [242] for degradation. These genes, in addition to the examples discussed throughout this section, hint towards patterns of nutrient availability which differ from the temperate Octadecabacter clade members. A concomitant difference in co-factor synthesis and utilisation was already indicated by the polar group-specific *mtnA* gene, and, in addition, a pyridoxamine 5'-phosphate oxidase gene is specific for polar Octadecabacters (octa_41850). This enzyme catalyses the last step in pyridoxal-phosphate (vitamin B_6) synthesis [243].

Finally, a copy of the small ribosomal subunit protein S21 (gene name *rpsU2*, octa_12240), specific to polar Octadecabacters, is of note: expression of this gene was found to increase drastically at lower temperatures in the Cyanobacterium *Anabaena variabilis* [32], and it is part of a cold shock-responsive operon found in *Sinorhizobium meliloti* [33]. Consequently, it seems to be directly involved in alleviating the effects of lower temperatures by stabilising the ribosomal complex, as opposed to countering indirect effects, such as increased oxidative stress.

3.4.2 *Genes specific to temperate Octadecabacter isolates*

The majority of genes specific to both *O. ascidiaceicola* and *O. temperatus* could not be assigned a specific function, as the products of

43 out of 68 specific genes are annotated as hypothetical proteins (compare Table A.3, p. 138). This indicates that the biochemical or metabolic features, which define this group, are generally poorly characterised. In addition, most of the functions assigned to group-specific genes are not exclusive to these two members. These genes constitute non-orthologous versions of functions, which are also part of the Octadecabacter core-genome (unlike with most of the polar Octadecabacter-specific genes discussed in the previous section).

Among the genes with an assigned function, those involved in stress resistance dominate numerically. However, none of them are organised in group-specific loci, as found in polar Octadecabacters. Two genes belong to the universal stress protein (Usp) A-family (COG 0589), namely the stress response protein NhaX (oasc_04570, a gene exclusively present in this group), and Usp F (oasc_20280). Its members are usually involved in the response to a wide range of stress signals, such as starvation, oxidative stress, or DNA damage [244]. For dealing with the latter, both Octadecabacters possess a group-specific copy each of DNA polymerase IV (oasc_15410), and a 3-methyl-adenine DNA glycosylase (oasc_25850). Like polar Octadecabacters, this group therefore seems to possess increased DNA repair capability, although both functions are also present in the Octadecabacter core-genome. The tellurite resistance protein TerB (oasc_12800) is specifically required to increase tellurite resistance, but is also likely involved in integrating more diverse stress signals [245]. Taken together, these group-specific genes hint towards differences in the general stress response network of this group. This may correlate with a group-specific orthologue of a toluene efflux pump precursor (TtgF, oasc_10420), as well as a poly-β-1,6-N-acetyl-D-glucosamine synthase (oasc_28310), and exopolysaccharide synthesis ExoD (oasc_02050), which are both involved in biofilm formation [246, 247]. Organic solvents as stress agents, and biofilm formation in response to stress signals may be of higher relevance for these particular strains, and could be integrated into the general stress response, among others, by some of the genes described above. Similar functionalities are also encoded in the Octadecabacter core-genome, and a detailed account of the environmental integration of this group would require experimental examination (e. g. via transcriptomic analyses). A further interesting characteristic is the presence of a group-specific spermidine synthase gene (oasc_20220), as this polyamine is involved in the protection of DNA against oxidative stress, and biofilm formation and surface-associated motility [248, 249]. A non-orthologous form of this enzyme is only present in the shell-genome of isolate NH9-P7.

Only few group-specific functions are associated with metabolism. They include an α-ketoglutaric semialdehyde dehydrogenase (oasc_21950), which catalyses the last step in L-arabinose, and L-proline degradation, among other substrates [250]. Thus, expansion of the substrate range beyond that encoded in the Octadecabacter core-genome is not

as pronounced in this group as it is in polar Octadecabacters, based on genes with functional annotation.

Notably, none of the functionally characterised, group-specific OGs are associated with genetic mobility (such as transposases or insertion sequences (ISs)), underlining that the abundance of such functions is a particular characteristic of polar Octadecabacters.

3.4.3 *Genes specific to Pseudooctadecabacter-related genomes*

The most striking characteristic of group III, formed by *P. jejudonensis* and isolate NH9-P7, is their genetic ability to perform aerobic anoxygenic photosynthesis (AAP), which constitutes 39 out of 78 group-specific genes (Table A.4, p. 140). Photochemical reaction centres generally seem to increase metabolic efficiency, i. e. they maximise the amount of biomass an organism can produce from its available nutrients [251]. Their activity generates radicals, and is therefore a source of oxidative stress [252]. Nonetheless, the former aspect seems to be more important in the ecology of the two members, underlined by the presence of a group-specific starvation-inducible DNA-binding protein (OJEJ_33070).

Two out of nine Octadecabacter clade strains being theoretically able to perform AAP fits the larger context of the *Roseobacter* group, for which 31 out of 116 strains possess the relevant genes.

Interestingly, despite a higher potential load of autogenic oxidative stress induced by performing AAP, the group-specific genetic equipment relating to the cellular stress response is small compared to groups I and II. For example, specific functions pertaining to DNA damage repair, or the protection of sulphur species are absent. Both strains possess a specific copy of a CspA-family β-ribbon cold-shock protein (OJEJ_11070). This family acts to some extend in the transcriptional regulation of genes involved in the cold-shock response, but, more importantly, likely halts translation upon cold-shock until the organism can adapt its metabolic network to the new conditions [30]. As such, this family is also present in all other Octadecabacter clade genomes, but seems to play a larger role in the ecology and metabolic regulation of *P. jejudonensis* and isolate NH9-P7.

As with group II, specific genes of group III are not organised in functional loci (except for AAP-genes), and do not contain transposases or other indicators of genetic mobility.

3.5 FUNCTIONAL GENE CONTENT EVOLUTION

Lineage-specific parameter values of the birth-and-death model are depicted in Figure 3.19 for an overview of each parameter's relative importance at a given node in the Octadecabacter clade. Note that all parameters are normalised to μ, which therefore equals 1 in all charts.

The plot in the upper right corner shows the correlation between the edge length parameter t estimated by COUNT, and the respective edge's length in the reference tree (Figure 3.2), subdivided into internal and terminal branches.

Changes in COGs at the polar Octadecabacters' latest common ancestor (LCA) are given in Table 3.5. Three polar *Loktanella* isolates were included in the birth-and-death model, which enabled a comparison of COG changes along their and the polar Octadecabacters' lineages. Only COG changes at internal, i. e. ancestral nodes were considered to that end, and are listed in Tables A.7 to A.11 (see Figure A.8 on p. 172 for an overview). The ancestor of the two temperate species *O. temperatus* and *O. ascidiaceicola* was included as a non-polar comparison group. In general, ancestors of polar strains do not show higher similarity to each other than to the temperate ancestor. This situation remains unchanged when gains and losses at their respective leaf nodes are included in the comparison. Two COGs were specifically newly gained at the ancestral lineages of all polar isolates: COG 1484 (DNA replication protein DnaC), and COG 2608 (Copper chaperone CopZ). Conversely, only one COG shows similar behaviour between the temperate and more than one polar ancestor: COG 0848 (Biopolymer transport protein ExbD).

General pathways affected by the COG content changes listed in Table 3.5 will be described in the following sections.

Table 3.5: Changes in COGs at the polar Octadecabacters' LCA, as predicted by COUNT. Only COGs with a probability of change $p > 0.5$ at this node are listed. The approximate probability ratios are given on the right, with green and red indicating gain and loss, respectively. Relevant probabilities are gain (G) : expansion (E) : neutral (N, no change), and loss (L) : reduction (R) : neutral (N). A box indicates the highest probability, and is filled if it surpassed the second-highest by at least 50 %. The rightmost column gives the orders of magnitude between the highest and lowest ratio (the latter is always 1). COGs are sorted by category, with the respective category code given on the left. COGs assigned to more than one category are listed multiple times.

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0578	Glycerol-3-phosphate dehydrogenase	1	75	73	10^3
C	1071	TPP-dependent pyruvate or acetoin dehydrogenase subunit α	1	51	27	10^4
	1359	Quinol monooxygenase YgiN	319	1	319	
	1454	Alcohol dehydrogenase, class IV	5	1	2	
	3794	Plastocyanin	3	4	1	
	5524	Bacteriorhodopsin	173	1	15	10^3

Table 3.5: continued

	COG ID	DESCRIPTION	G:E:N	L:R:N		
E	0182	Methylthioribose-1-phosphate isomerase (methionine salvage pathway), a paralog of eIF-2B α subunit	105	1	19	10^3
	0263	Glutamate 5-kinase	1	4	1	10^5
	0620	Methionine synthase II (cobalamin-independent)	497	1	200	
	1231	Monoamine oxidase	1	11	5	
	2113	ABC-type proline/glycine betaine transport system, periplasmic component	1	55	40	10^4
	2986	Histidine ammonia-lyase	39	1	14	10^2
	2987	Urocanate hydratase	44	1	15	10^2
	3741	N-formylglutamate amidohydrolase	1	50	17	10^4
	3931	Predicted N-formylglutamate amidohydrolase	1	36	22	10^3
	4175	ABC-type proline/glycine betaine transport system, ATPase component	1	38	27	10^4
	4176	ABC-type proline/glycine betaine transport system, permease component	1	38	27	10^4
F	0737	2′,3′-cyclic-nucleotide 2′-phosphodiesterase / 5′- or 3′-nucleotidase, 5′-nucleotidase family	1	4	3	10^3
	1051	ADP-ribose pyrophosphatase YjhB, NUDIX family	1	30	9	10^3
G	0058	Glucan phosphorylase	78	1	7	10^3
	1640	4-α-glucanotransferase	78	1	7	10^3
	2376	Dihydroxyacetone kinase	745	1	465	
H	0351	Hydroxymethylpyrimidine / phosphomethylpyrimidine kinase	25	1	1	
	0352	Thiamine monophosphate synthase	1	95	16	10^2

Table 3.5: continued

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0414	Pantothenate synthetase	184	1	28	10^3
	0819	Thiaminase	116	1	51	
	2022	Thiamin biosynthesis thiazole synthase ThiGH, ThiG subunit	1054	1	235	
	2145	Hydroxyethylthiazole kinase, sugar kinase family	3009	1	631	
	2241	Precorrin-6B methylase 1	26	1	18	10^2
	2242	Precorrin-6B methylase 2	41	1	11	10^3
K	1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)	1	34	18	10^5
	4567	Two-component response regulator, ActR/RegA family, consists of REC and Fis-type HTH domains	1	32	5	10^4
L	0582	Integrase	1	8	8	
	1484	DNA replication protein DnaC	531	1	40	
	3593	Predicted ATP-dependent endonuclease of the OLD family, contains P-loop ATPase and TOPRIM domains	4740	1	1053	
	3598	RecA-family ATPase	223	1	59	
M	0381	UDP-N-acetylglucosamine 2-epimerase	5122	1	2379	
O	0694	Fe-S cluster biogenesis protein NfuA, 4Fe-4S-binding domain	1	153	30	10^4
	1305	Transglutaminase-like enzyme, putative cysteine protease	1	1	2	
	2761	Predicted dithiol-disulfide isomerase, DsbA family	1	38	16	10^3
P	0025	NhaP-type Na^+/H^+ or K^+/H^+ antiporter	1	29	22	
	2608	Copper chaperone CopZ	3852	1	1914	

Table 3.5: continued

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	4638	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, large terminal subunit	1	15	10	10^3
Q	1228	Imidazolonepropionase or related amidohydrolase	107	1	67	
R	0385	Predicted Na^+-dependent transporter	652	1	644	
	2321	Predicted metalloprotease	32	1	24	10^2
	2910	Putative NADH-flavin reductase	1507	1	569	
	3380	Predicted NAD/FAD-dependent oxidoreductase	105	1	34	10^2
	3450	Predicted enzyme of the cupin superfamily	932	1	87	
	3607	Predicted lactoylglutathione lyase	61	1	41	10^2
	4122	Predicted O-methyltransferase YrrM	41	1	3	10^3
	4638	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, large terminal subunit	1	150	98	10^2
S	0398	Uncharacterized membrane protein YdjX, TVP38/TMEM64 family, SNARE-associated domain	1	88	33	10^3
	1357	Uncharacterized protein YjbI, contains pentapeptide repeats	72	1	32	
	2833	Uncharacterized conserved protein, contains ferritin-like DUF455 domain	24	1	4	10^3
	2841	Uncharacterized conserved protein YdcH, DUF465 family	57	1	18	10^3
	4067	Uncharacterized conserved protein	75	1	10	10^3
	5345	Uncharacterized protein	303	1	10	10^3

Table 3.5: continued

	COG ID	DESCRIPTION	G:E:N	L:R:N		
T	1409	3′,5′-cyclic AMP phosphodiesterase CpdA	4	4	1	
	1974	SOS-response transcriptional repressor LexA (RecA-mediated autopeptidase)	1	340	176	10^4
	4567	Two-component response regulator, ActR/RegA family, consists of REC and Fis-type HTH domains	1	320	54	10^3
	5524	Bacteriorhodopsin	173	1	15	10^3
U	3505	Type IV secretory pathway, VirD4 component, TraG/TraD family ATPase	1196	1	360	
V	0737	2′,3′-cyclic-nucleotide 2′-phosphodiesterase/ 5′- or 3′-nucleotidase, 5′-nucleotidase family	1	4113	3321	
	1787	Endonuclease, HJR/Mrr/RecB family	2075	1	1765	
	2746	Aminoglycoside N3′-acetyltransferase	462	1	37	10^3
X	0582	Integrase	1	8	8	
	1943	REP element-mobilizing transposase RayT	2	1	1	
	2826	Transposase and inactivated derivatives, IS30 family	237	22	1	
	3415	Transposase	970	1	619	
	3547	Transposase	18	6	1	
	5433	Predicted transposase YbfD/YdcC associated with H repeats	1841	1	983	

3.5.1 *Thiamine metabolism*

All sequenced Octadecabacter clade members possess the genes necessary for thiamine uptake (via the ABC-transporter ThiBPQ) and subsequent conversion to thiamine di- and triphosphate (via the thiamine

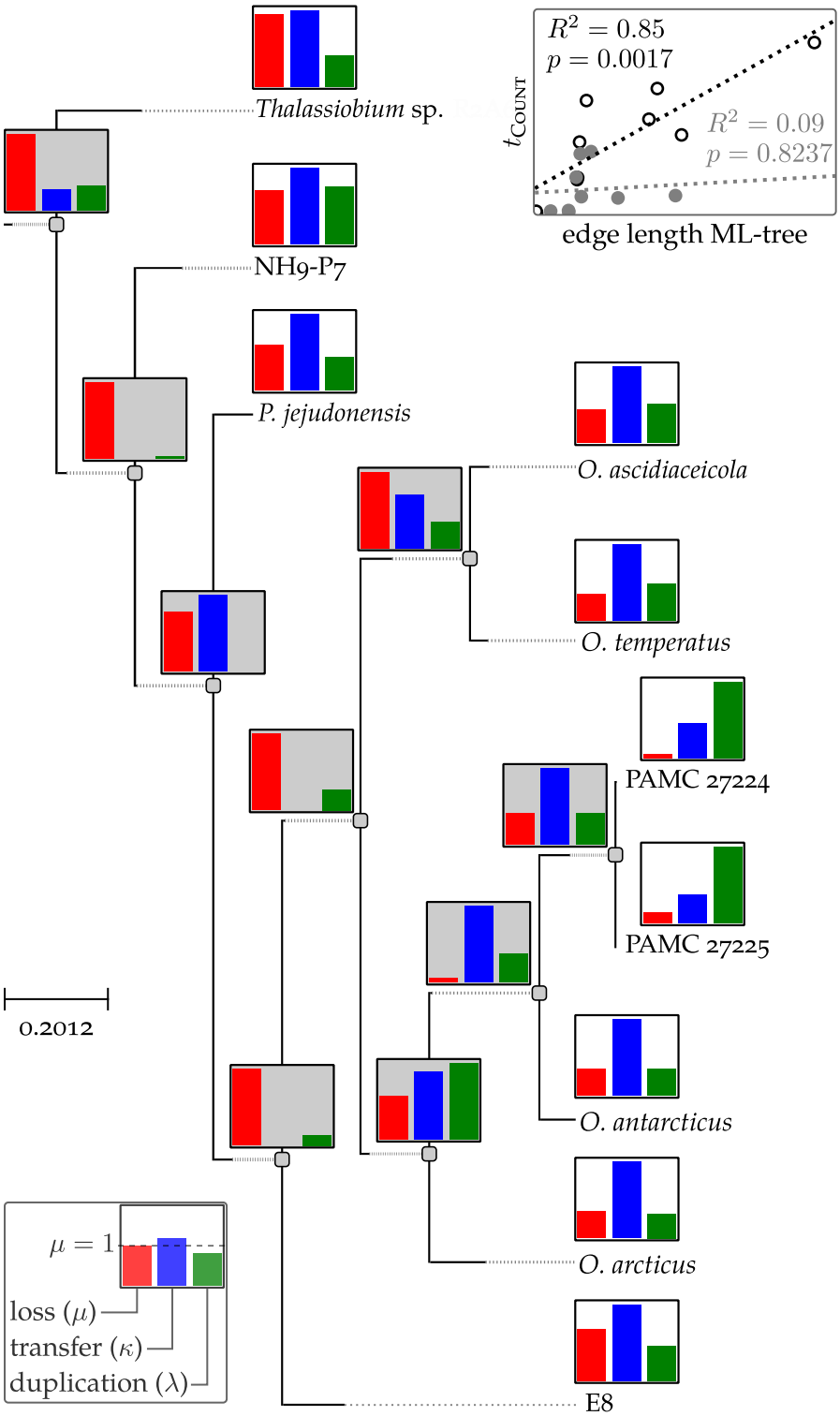


Figure 3.19: Lineage-specific components of transfer (κ), loss (μ), duplication (λ), and edge length (t) parameters, as estimated by COUNT (Section 2.9). Tree presents t directly (but disregard dotted lines), while the other parameters are provided in bar charts. All parameters (including t) are normalised to μ, which therefore equals 1 in all charts. Charts at internal nodes are filled with grey background. Plot shows relation between t and edge length in the ML-tree for internal (grey) and terminal nodes (hollow circles).

pyrophosphatase ThiN). In addition, the polar Octadecabacters, as well as *O. ascidiaceicola*, can synthesise thiamine de novo: COGs 0351, 0352, and 2022 constitute three genes in thiamine biosynthesis (thiD, E, and G, respectively), to which the phylogenetic birth-and-death model assigns high likelihoods of gain (thiD, G), or expansion (thiE) at the polar Octadecabacters' ancestral node (see Table 3.5, category H). There are multiple other genes required for thiamine synthesis (summarised well by Jurgenson, Begley, and Ealick [253]), all of which are present in the five genomes. The model assigns a higher probability to the independent acquisition of the respective genes by *O. ascidiaceicola*, than to them being present in a common ancestor of (and subsequent vertical heredity to) the polar Octadecabacters and *O. ascidiaceicola* (and their loss in *O. temperatus*). Interestingly, the five genomes also contain a single copy of a THI5-like gene. These are known from yeasts, in which they synthesise hydroxymethyl pyrimidine (HMP), one of the two thiamine precursors, from pyridoxin and histidine [254]. Usually, in prokaryotes ThiC, which is also present in all five Octadecabacter clade genomes, synthesises this compound from 5-aminoimidazole ribotide, an intermediate in purine biosynthesis [253]. In contrast to THI5, ThiC consumes S-adenosylmethionine (SAM) (in form of a 5´-deoxyadenosyl radical), and requires a 4Fe-4S cluster as co-factor.

COGs 0819 (TenA) and 2145 (ThiM) encode enzymes involved in thiamine salvage [253], the latter being exclusive to the polar members. Of the two, TenA cleaves the unphosphorylated form of thiamine and seems to function in recycling of the HMP moiety [255], while ThiM functions in thiazole-salvage [253].

Generally, thiamine, like the other B vitamins, is a co-factor often exchanged between microbes in mutualistic relationships [256]. Presence of its biosynthetic pathway in the polar Octadecabacters and *O. ascidiaceicola* therefore reflects differences in these relationships in their respective environments compared to the other clade members. Thiamine is produced by the majority of marine prokaryotes, while many marine eukaryotes are auxotrophic for it [256]. The latter comprise a large fraction of the active microbes in sea ice [41], and consequently, polar Octadecabacters may take up the role of thiamine providers in exchange for, e. g. photosynthates, a relationship which was demonstrated for other *Roseobacter* group bacteria as well [257]. An alternative explanation may be that the Octadecabacters preferentially take up thiamine from the environment, and only synthesise it during phases of low nutrient availability. Polar Octadecabacters are, for example, confronted with such situations when the sea ice melts, and they are expelled to the open ocean. Their need to efficiently recycle nutrients in order to survive is further signified by the fact that they possess genes for thiamine degradation which are absent in the rest of the clade.

3.5.2 *Tetrapyrrole biosynthesis*

All sequenced Octadecabacter clade genomes encode genes for the synthesis of the most common tetrapyrroles heme, siroheme, and cobalamin. The common precursor δ-aminolevulinic acid is synthesised from glycine and succinyl-CoA by the gene product of *hemA* in the C4-pathway, which is prevalent in α-Proteobacteria [234]. A notable exception exists in the two members *P. jejudonensis* and isolate NH9-P7, which are also genetically equipped to synthesise bacteriochlorophyll and perform AAP (see Section 3.4.3).

Orthologues to one of the many methyltransferases in cobalamin biosynthesis, precorrin-6B methylase (*cobL*), were assigned to different COGs in different Octadecabacters, and the phylogenetic birth-and-death model predicted their respective gain and loss at the polar Octadecabacters' LCA (COGs 2241 and 2242 in Table 3.5, category H). Notably, both COGs were registered as orthologues in the present analysis. Nonetheless, they differ in length (≈240 AA and ≈200 AA for COG 2241 and 2242, respectively) and show different alignment patterns within the COG reference organisms. The polar Octadecabacters seem to have lost the second isoform (COG 2422) in favour of the first one (COG 2421). Among the non-polar Octadecabacters, the only other occurrence of this enzyme is in isolate NH9-P7. Within the whole *Roseobacter* group, COG 2422 seems to be the more prevalent, as it is shared by 85 of the 115 analysed genomes (compared to 29 for COG 2421). No clear distinction in localisation of the two enzymes seems to exist, based on the isolation coordinates of the *Roseobacter* group members (Table A.1). Neither does life style (as estimated from the isolation circumstances) serve as a predictor for the isoform ($\chi^2 = 7.03$, $p = 0.32$ and $\chi^2 = 3.86$, $p = 0.7$ for COGs 2241 and 2242, respectively).

Consequently, a slight difference seems to exist between polar and non-polar Octadecabacter clade members concerning cobalamin synthesis.

3.5.3 *Methionine metabolism*

All Octadecabacters possess the necessary genes for methionine synthesis, with both methylmethionine and methyl-tetrahydrofolate (MTHF) as methyl donors. The genes likely constitute the homocysteine-responsive *metR* regulon specific for *Rhodobacterales*, as recently described by Leyn et al. [258]. In addition to the cobalamin-dependent methionine synthase (*metH*), the polar Octadecabacters seem to have gained the cobalamin-independent synthase (*metE*, see Table 3.5). This enzyme transfers the methyl group from MTHF to homocysteine via a mechanism independent from cobalamin, albeit at a markedly lower rate [259]. All Octadecabacters are able to produce cobalamin, as described in Section 3.5.2. However, its production is costly, and the cobalamin-

independent methionine synthase may provide an alternative for the polar Octadecabacters during phases of low nutrient availability.

A further characteristic of polar Octadecabacters is the *mtnA* gene (coding for methylthioribose-1-phosphate isomerase; see Table A.2, and COG 0182 in Table 3.5), the product of which catalyses the second step of the methionine salvage pathway [260]. This pathway recycles methylthioadenosine (MTA), a product of SAM utilisation, to regain methionine. MTA is produced through donation of an aminopropyl group from SAM, e. g. in polyamine or N-acetyl-homoserine lactone synthesis. The first step of the methionine salvage pathway is phosphorolysis of MTA, which is catalysed by the product of *mtnP*, present in all Octadecabacters. In this step, adenine and methylthioribose (MTR)-1-phosphate are formed. The latter is then isomerised to methylthioribulose-1-phosphate by the *mtnA*-coded isomerase. Subsequent conversion to methionine happens in five consecutive steps, catalysed by a dehydratase, enolase, phosphatase, dioxygenase and transaminase [261]. Notably, almost none of the respective genes could be found in the polar Octadecabacters, neither by their annotation, nor through KO terms. Only one homologue of the enolase seems to be coded by the PAMC isolates. The enolases usually involved in this reaction belong to the haloacid dehalogenase superfamily [261]. While the polar Octadecabacters have, on average, more genes of this superfamily than the temperates, there is no enolase specific for this group. Interestingly, RuBisCO-like proteins were also found to catalyse this reaction [261], and one of the genes in *O. antarcticus* is annotated with this function [61]. However, it does not occur as orthologue in any other Octadecabacters. The dioxygenase is usually of the cupin-superfamily [261], and there is a high probability that one such gene was exclusively gained by the polar Octadecabacters (COG 3450 in Table 3.5, category R). A high degree of flexibility is also known for the last step of the cycle, transamination to methionine [261].

Generally, the methionine salvage pathway conserves sulphur within the cell. It is only of importance, if larger amounts of SAM are used in polyamine or homoserine lactone synthesis. The polar Octadecabacters lack polyamine synthases, and therefore likely do not produce large amounts of MTA. From the produced MTA, adenine is universally recycled in the Octadecabacters as described above, but instead of recycling the resulting MTR-1-phosphate, they likely export it. A similar behaviour was observed in *E. coli* [262]. In *E. coli*, absence of the methionine salvage pathway was attributed to the high sulphur availability in its natural environment, which may also exist for the Octadecabacters (discussed in Section 4.3.2). MTR-1-phosphate has no known immediate use for cellular biochemistry, but may serve as quorum sensor in the extra-cellular space [262]. However, the question why the polar Octadecabacters possess the *mtnA* gene remains.

3.5.4 *Histidine degradation*

All polar Octadecabacters, as well as isolate E8 are able to use histidine as carbon and nitrogen source. COGs 2986 and 2987, which were likely independently gained by the polar Octadecabacters' LCA (Table 3.5), are the first two steps in histidine degradation, a pathway discussed in detail in ref. [263]. Histidine ammonia-lyase (*hutH*) deaminates histidine, producing ammonium and urocanate. The latter is then hydrated to imidazolonepropionate (*hutU*, COG 2987 in Table 3.5), and further hydrolysed to formiminoglutamate by an imidazolonepropionase (HutI, COG 1228 in Table 3.5, category Q). In Octadecabacters, the *hutF* gene product (formiminoglutamase/formiminoglutamate deiminase) cleaves a further ammonium molecule off the formimino-group, leaving formylglutamate. In a last step, N-formylglutamate amidohydrolase or formylase, coded for by *hutG*, cleaves formylglutamate to formate and glutamate. While a gene encoding this enzyme is present in all Octadecabacters, the polar species possess more copies, and expansion likely occurred at their LCA (see COGs 3741 and 3931 in Table 3.5). Note that the ABC-type proline/glycine betaine system, which received an expansion in the polar lineage (COGs 2113, 4175 and 4176) may as well code for histidine uptake transporter, as the two are regularly confused [263].

3.5.5 *General cellular stress response*

Genes potentially involved in the response to increased levels of oxidative, osmotic, or respiratory stress comprise a significant portion of the polar Octadecabacter-specific genes discussed in Section 3.4.1. Fittingly, similar functions were gained or expanded at the polar Octadecabacters' LCA, some of which are identical to those found group-specific (e. g. NfuA).

Among the COGs gained, some are associated with a chaperone function, such as COG 2608, which represents the copper chaperone CopZ. CopZ binds copper ions for their export, and thus prevents their exposition to the cellular environment, which would result in oxidative stress [264]. Interestingly, this COG was likely gained in the ancestral lineages of all polar isolates contained in the birth-and-death model, including polar Loktanella species (Section 3.5). In addition, a DsbA-family dithiol-disulphide isomerase (COG 2761) likely experienced family expansion at this node. This family mediates the formation of disulphide bonds in proteins exported to the periplasm [265], and thus ensures their proper structure and function.

The Octadecabacter LCA furthermore likely acquired a quinol monooxygenase (COG 1359). This enzyme oxidises quinone-derived substrates, and therefore plays a role in maintaining the proper balance of quinone and quinol pools in the electron transport chain [266]. No-

tably, it does not require co-factors. In addition, the functional family of plastocyanins (COG 3794) likely experienced family expansion at this point, in accordance with the fact that one such protein is specific for the polar group (annotated as pseudoazurin, Table A.2). These copper-binding proteins are usually involved in electron transfer reactions to cytochromes in diverse contexts [267]. Gain and expansion of these two functions likely serves polar Octadecabacters in oxidative or respiratory stress protection.

A possible expansion of NhaP-type cation/H^+ antiporters (COG 0025) may complement this functionality, since these are usually involved in maintenance of the intracellular pH [268]. In combination with the polar Octadecabacters' increased capacity for proton-gradient generation (due to their possible xanthorhodopsin phototrophy), this transporter type may also simply serve as a sodium exporter, and thus function to counteract osmotic stress. In addition, glutamate kinase (COG 0263), may have likely expanded at the polar Octadecabacter LCA. This enzyme catalyses the first step in proline synthesis from glutamate [269]. Proline, as well as proline betaine are both widespread osmoprotectants [270, 271]. Additional copies of this gene may serve to compensate for its feedback-inhibition by proline itself, and thus allow for the accumulation of larger proline pools during phases of osmotic stress.

Furthermore, ADP-ribose pyrophosphatase (COG 1051) likely expanded along the polar Octadecabacters' lineage. This enzyme cleaves ADP-ribose into AMP and ribose-5-phosphate, but most of its studied homologues also show reduced activity on ADP-sugars and nicotinamide adenine dinucleotide (NAD) [272, 273]. ADP-ribose is a product of NAD turnover and has potential cytotoxic effects, most importantly through non-enzymatic protein glycation [274]. Consequently, expansion in this gene family may reduce cellular damage by this compound. In addition, this class of pyrophosphatases also contains the tellurite resistance protein TrgB, which is a major player of tellurite resistance in *E. coli* and *Rhodobacter capsulatus* [272, 275], and in the latter was shown to also increase resistance to copper [275].

Some transcriptional regulators associated with different types of stress are specific for polar Octadecabacters (Section 3.4.1). Generally, DNA damage (induced either by UV-irradiation, or oxidative stress) induces a stress-responsive network of genes in what is referred to as the SOS-response [276]. This response is mainly regulated by two proteins: LexA acts as transcriptional repressor of SOS-response-associated genes, and, upon accumulation of DNA damage, its auto-catalytic cleavage is induced by RecA [276]. The repressor LexA experienced gene family expansion with high probability in the polar Octadecabacters' ancestral lineage (COG 1974 in Table 3.5). This indicates a tighter control of the SOS-response in polar Octadecabacters, which concurs with their altered genetic equipment dealing with DNA damage (DNA polymerase V).

In summary, the diverse stress-related functions discussed above supplement those found among the polar Octadecabacter group-specific genes.

3.5.6 *Other metabolism-related functions*

Apart from (more or less) complete metabolic pathways probably gained by the polar Octadecabacters' LCA, which were described in previous sections, some single COGs showed change at this point in the phylogeny, which are associated with individual metabolic reactions and stages. One of these is COG 1409 (cyclic adenosine monophosphate (AMP) (cAMP) phosphodiesterase CpdA), which degrades cAMP and thus regulates cellular pools of this important effector molecule. In effect, *cpdA*-overexpressing *E. coli* cells showed increased resistance to oxidative stress, mediated by a strong induction of *rpoS* expression [277]. Independent from this function, CpdA also seems to be involved in global regulation of amino acid synthesis [278].

Notably, another cyclic nucleotide phosphodiesterase (COG 0737) is among the few functions that experienced family reduction along the polar Octadecabacter lineage. This periplasmic enzyme degrades cyclic nucleotides with broad specificity [279]. As such, it is important in utilising extracellular DNA as a carbon and phosphate source [280], and may be particularly relevant in habitats with phosphate-limitation. Moreover, Trülzsch et al. observed that this enzyme mediates the ability of *Yersinia enterocolitica* to grow on 2´,3´-cAMP as sole carbon and energy source [281].

Polar Octadecabacter-specific genes indicated a differential utilisation of glycerol and its intermediates in this group (Section 3.4.1). In addition to gaining the group-specific DHA kinase, the polar Octadecabacter LCA likely expanded in glycerol-3-phosphate dehydrogenase (GPDH) (COG 0578) functionality. This enzyme is part of the glycerol degradation pathway found in all Octadecabacters. Its increased numbers in polar representatives may reflect higher availability of this compatible solute in their habitat, due to its production by eukaryotes upon osmotic stress [209, 282]. Notably, GPDH links the fatty acid cycle with glycolysis via its inter-conversion of DHAP and glycerol-3-phosphate. Expansion of this gene family may therefore serve a regulatory function, being a further indicator that polar Octadecabacters possess more complex metabolic networks, and need to integrate more complex environmental stimuli.

Apart from glycerol, primary producers frequently secrete storage saccharides, which may be degraded and metabolised by associated prokaryotes [283]. The polar Octadecabacters' LCA likely gained the group-specific 4-α-glucanotransferase (COG 1640), as well as a glucan phosphorylase (COG 0058); both are enzymes for the mobilisation of these polysaccharides.

Finally, polar Octadecabacters seem to have lost the ability to synthesise pantothenate at their LCA (loss of pantothenate synthetase PanC, COG 0414). Pantothenate is the precursor of CoA, and thus essential for cellular metabolism [284]. However, it is also produced in excess quantity and secreted by many bacteria [284], and may thus be easily taken up from the environment by pantothenate-auxotrophs.

3.6 PARALOGUE OCCURRENCE IN OCTADECABACTER CLADE MEMBERS

Paralogues are derived from gene duplications within an organism, and therefore show highest BLAST-based similarity to their original gene in the same genome [285]. To test the amount of duplications present in the Octadecabacter genomes, determination of paralogous relationships was included in the orthology detection (Section 2.5).

Generally, polar Octadecabacters possess more paralogues than the temperate strains, ranging from an average of 68 in the PAMC isolates to over 560 in *O. arcticus*, versus 22 to 45 in the other strains (Table 3.6). Isolate NH9-P7 constitutes a notable exception, with over 150 paralogues in total. While most duplications in the polar type strains of *O. arcticus* and *O. antarcticus* are attributable to mobile genetic elements, i. e. transposases, integrases, and ISs, they still possess a higher average number of paralogues than the others when these cases are excluded (Table 3.6).

The relative paralogue content of polar Octadecabacters, normalised to the number of CDSs (as this group possesses on average ≈800 CDSs more than the temperate strains, Table 3.1), is still high compared to the other groups. It ranges from an average of 13.8 paralogues in different OGs per 1,000 CDSs in the PAMC isolates to 31.6 in *O. arcticus* and *O. antarcticus* (excluding mobile genetic elements). Temperate strains possess an average of 9 paralogues in different OGs per 1,000 CDSs, with the exception of isolate NH9-P7 (44.2). Notably, the high fragmentation of the PAMC-isolate genomes (Table 3.1) could lead to an underestimation of their total paralogue content (see Appendix A.7, p. 203).

The majority of paralogues (45 to 92 %) is strain-specific, only the highly similar PAMC isolates pose an exception. There are no paralogues which are specific to the polar Octadecabacters (group I), i. e. which are present in all members of this group, and absent in all members of groups II to IV. All polar isolates do, however, possess multiple paralogues in the group-specific genome (as listed in Table A.2). In fact, the fraction of their group-specific genes being duplicated in at least one of their genomes is larger than that of the other groups.

Differences are also visible in the distribution of paralogues among the core-, cloud-, and shell-genomes defined in Section 3.3.1. While gene duplications in the type strains of *O. arcticus* and *O. antarcticus* predominantly affect their shell-genomes, isolate NH9-P7, with its similar

Table 3.6: Number of paralogues in Octadecabacter clade genomes. Grey numbers exclude the strain's mobile elements[a]. Group-specific paralogues are present in all members of a group (as defined in Section 3.2), and absent in all other strains. GROUP-SPECIFIC GENES as listed in Appendix A.4. Core-, cloud-, and shell-genome boundaries as defined in Section 3.3.1.

	PARALOGUES...		SPECIFIC FOR...		PARALOGUES IN...	
	IN OGS	TOTAL	STRAIN	GROUP	GROUP-SPECIFIC GENES	CORE-, CLOUD-, & SHELL-GENOME
O. arcticus	371 (162)	568 (192)	342	0	29	
O. antarcticus	220 (130)	303 (138)	186			
PAMC 27224	64 (54)	67 (56)	14			
PAMC 27225	67 (61)	69 (63)	14			
O. ascidiaceicola	22 (22)	22 (22)	10	0	2	
O. temperatus	35 (34)	35 (34)	25			
P. jejudonensis	25 (24)	25 (24)	15	1	3	
NH9-P7	154 (152)	155 (153)	137			
E8	44 (40)	45 (41)	32	32	20	

[a] all OGs whose annotations contained the strings "obile_element", "ransposase", "ntegrase", or "IS"

number of paralogues, shows more duplications in the Octadecabacter clade core-genome (Table 3.6).

3.7 KEY CHARACTERISTICS IN POLAR ROSEOBACTERS

The comparison genomes used in this study included five polar isolates, which were associated with clades other than the Octadecabacters (Table 3.7). Key figures, differing between polar and temperate Octadecabacters, were compared to these polar isolates and their associated clade members (Figure 3.20).

While polar Octadecabacters possess larger genomes with more protein-coding genes than their temperate relatives, this trend is not observable in the comparison groups. Regarding singletons, all polar isolates lie above their respective group's median (except for the highly similar PAMC isolates), but only the Octadecabacters possess the most

Table 3.7: Isolates outside of the Octadecabacter clade, which were procured from polar environments based on their isolation coordinates (Table A.1), are listed together with the *Roseobacter* group bacteria with which they form common clades (according to the MLSA phylogeny, Figure A.2). These were used for comparison of key figures between polar and non-polar isolates throughout the *Roseobacter* group (Figure 3.20).

POLAR ISOLATES	ASS. CLADE MEMBERS
L. vestfoldensis DSM 16212, *L. fryxellensis* DSM 16213, *L. salsilacus* DSM 16199	*Loktanella* genus-associated genomes in Table 3.2
Pseudophaeobacter arcticus DSM 23566	*Leisingera* genus-associated genomes in Table 3.2
Sulfitobacter guttiformis KCTC 32187	*Sulfitobacter* genus-associated genomes in Table 3.2, plus *S. donghicola* JCM 14565, *Sulfitobacter* sp. EE-36, *Sulfitobacter* sp. NAS-14.1, *Sulfitobacter* sp. CB2047

within their clade. Furthermore, in three of the four analysed clades, polar isolates show the highest number of paralogues, with the polar *S. guttiformis* as the only exception. The differences in the Loktanella and Leisingera clades are, however, less pronounced than in the Octadecabacters.

The polar Octadecabacters' capacity for carbohydrate uptake and metabolism is likely higher than that of the temperate ones, as they possess more genes associated with COG category G. Among the other groups, only polar *Loktanella* species mirror this trend to a degree (2 out of 3 polar isolates). In addition, polar isolates of both clades concordantly possess more genes in functions associated with recombination and genetic mobility (COG categories L and X, respectively), whereas such a difference is lacking in the other two comparison groups.

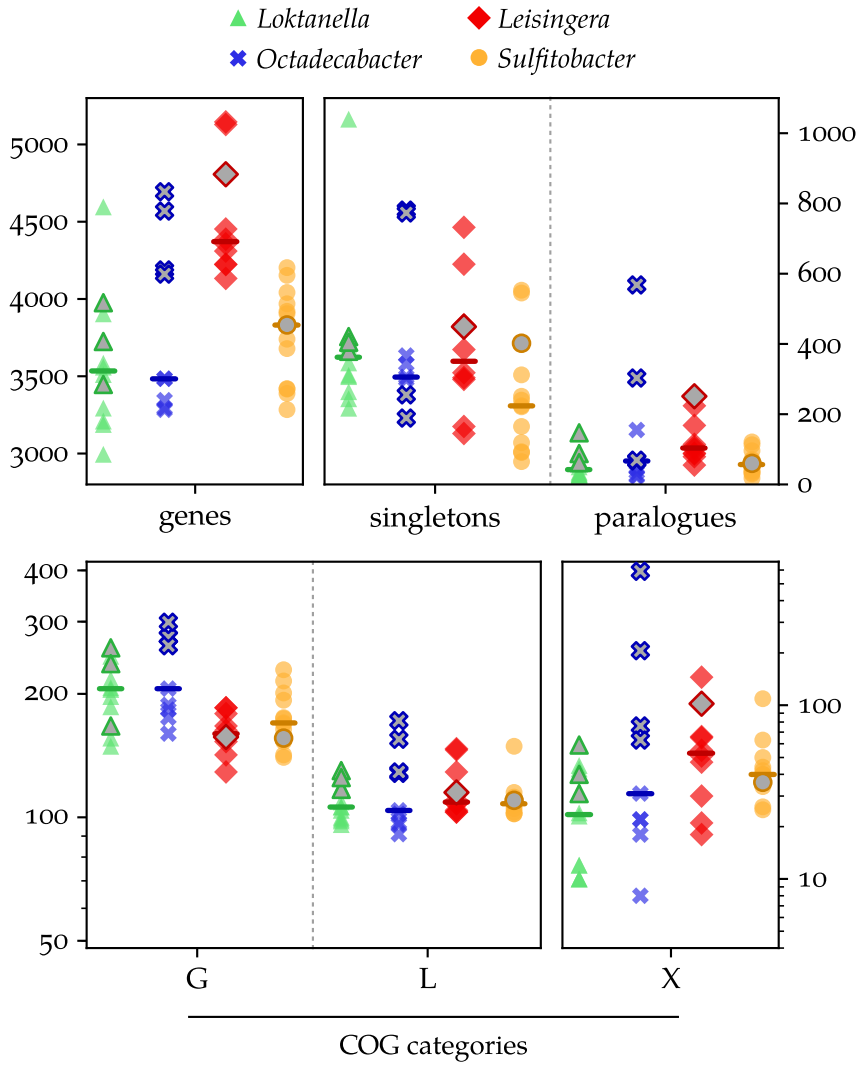


Figure 3.20: Key figures in polar roseobacters. Characteristics found to differ between polar and temperate Octadecabacters were compared to other *Roseobacter* group clades which harbour polar isolates. Their predominant associated genera are given on the top. Larger symbols filled with grey represent the polar isolates. Lines present median values. *genes* – number of protein-coding genes; *singletons* relating to all comparison genomes (Table 2.2)

3.8 PROTEIN SECONDARY STRUCTURE CONTENT

Secondary structure content of all proteins coded in the Octadecabacter genomes was estimated using the `jpred4` algorithm (Section 2.11). Pairwise euclidean distances in predicted α-helix and β-sheet content between the Octadecabacter core proteins were used for hierarchical clustering. When both structural features are considered together, polar Octadecabacters cluster separate from the rest, and groups defined in Section 3.2 are coherent (Figure 3.21). However, this coherence is absent when both features are considered separately.

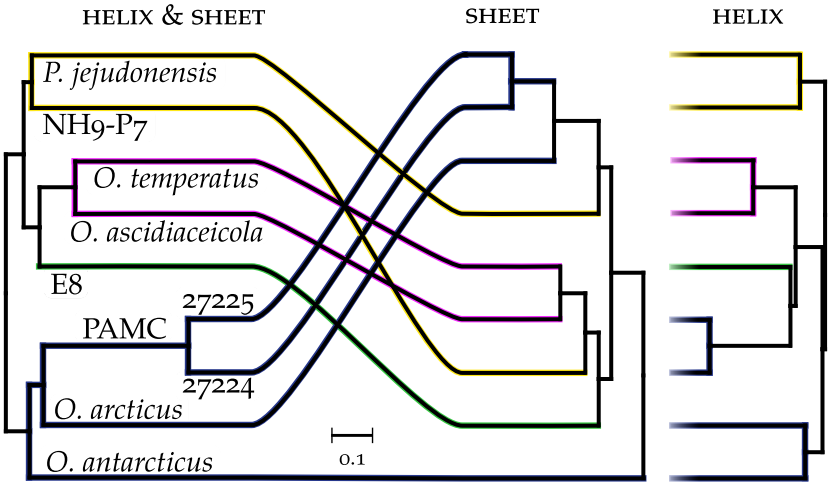


Figure 3.21: Hierarchical clustering of Octadecabacters based on predicted protein secondary structure. Clustering was based on the pairwise euclidean distances in relative secondary structure content between all Octadecabacter clade core proteins. Both structural features (β-sheet and α-helix) were considered in combination (left), as well as individually (centre and right, respectively). Colours mark groups defined in Section 3.2.

Wilcoxon signed-rank tests revealed that no Octadecabacter group shows a clear and statistically significant tendency towards containing different amounts of either structural feature than the others (Figure 3.22). Solely *O. antarcticus* possesses significantly less predicted β-sheets when compared to the members of group III. A similar difference in its α-helix content can only be observed for isolate NH9-P7 (and not for *P. jejudonensis*).

The performed χ^2 tests detected significant ($p < 0.05$) differences in observed versus expected amino acid count for the seven compared Octadecabacter strains. However, this significance likely stems from the high number of observations, as the mean differences in amino acid frequencies between the groups are very small (on the order of 10^{-3}, Figure 3.23). Nonetheless, polar Octadecabacters show some clear trends. For example, they possess higher portions of the hydrophobic amino acids alanine, isoleucine, and methionine (single letter codes A, I, and M, respectively). Within negatively charged amino

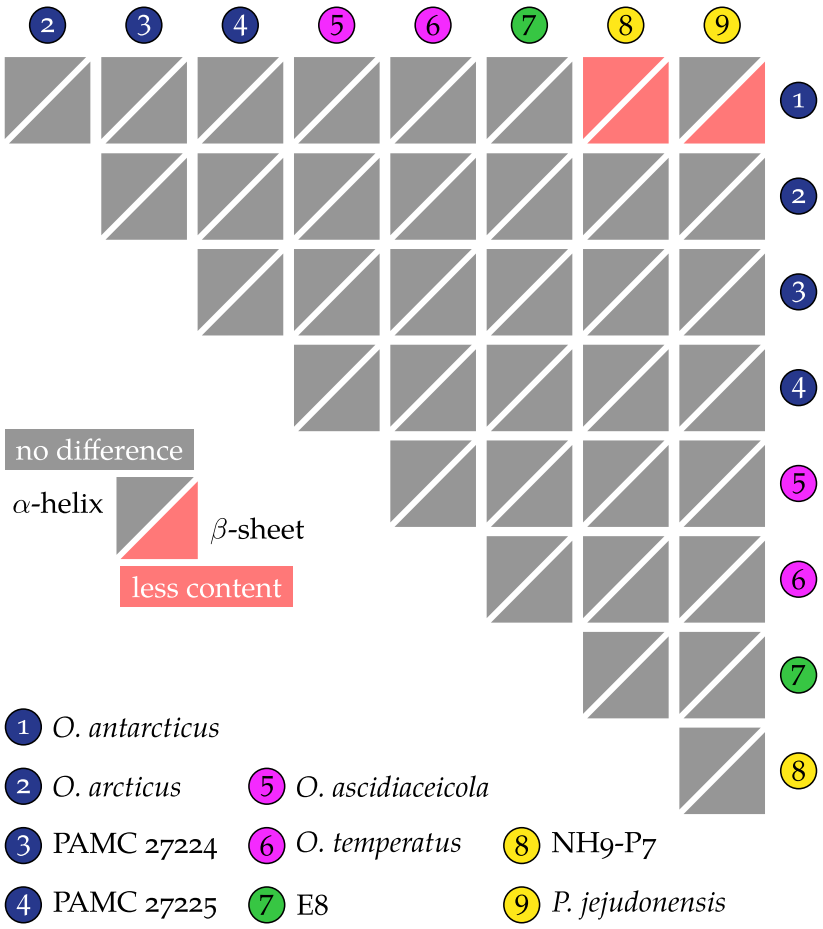


Figure 3.22: Results of Wilcoxon signed-rank tests for significant difference in secondary structure content between Octadecabacter clade members. Only *O. antarcticus*, when tested against the members of group III (as defined in Section 3.2), showed significantly different predicted contents under the Benjamini-Hochberg-corrected significance levels (Section 2.11). The `jpred4` algorithm predicts overall less β-sheet in this strain's core-proteome than in both isolate NH9-P7 and *P. jejudonensis*, and less α-helix than in the former (marked in red).

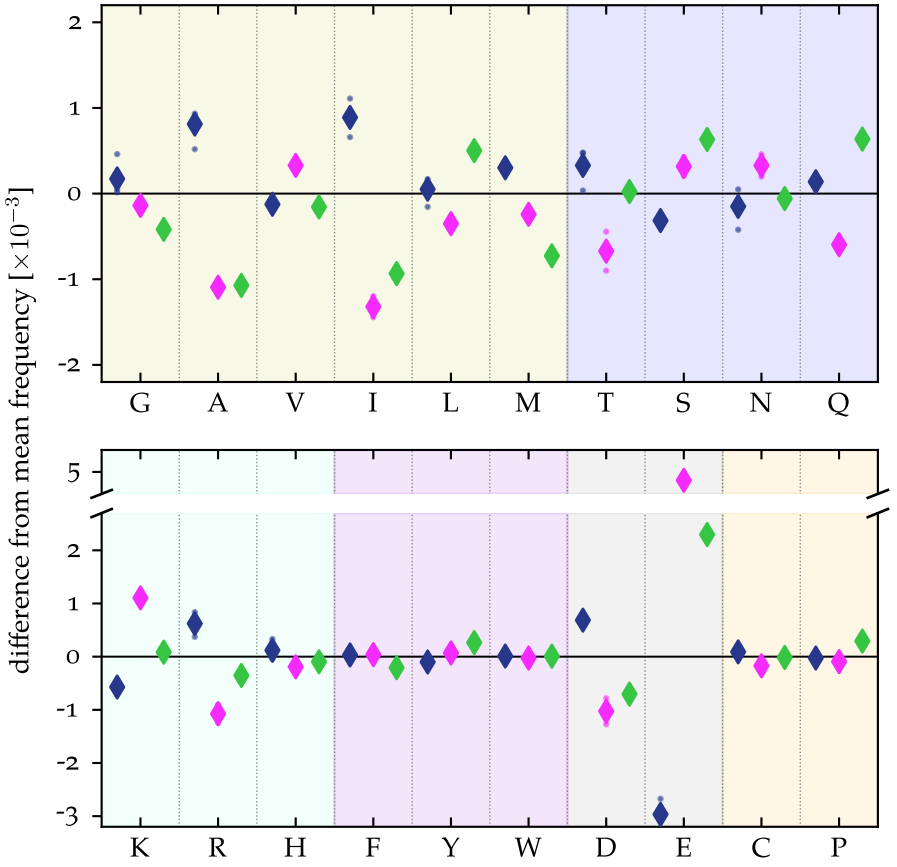


Figure 3.23: Mean amino acid frequency deviations in Octadecabacter groups. The mean amino acid frequencies in the Octadecabacter core-proteome were calculated for seven strains with similar GC-content (see Table 3.1), which included groups I, II, and IV (as defined in Section 3.2). The plot shows the difference of each group's mean to the base frequency, i. e. the mean frequency over all seven strains. Amino acids are sorted into hydrophobic (G, A, V, I, L, M), polar uncharged (T, S, N, Q), aromatic (F, Y, W), and negatively (D, E) and positively (R, H, K) charged groups.

acids, the group seems to favour aspartic acid (D) over glutamic acid (E). Finally, the four strains contain slightly more arginine (R), and slightly less serine (S) and lysine (K) than groups II and IV (Figure 3.23).

Compared to their general frequencies, the ratio of amino acids in predicted helix versus non-helix regions shows a higher within-group variation (compare individual values in Figure 3.23 versus Figure 3.24). Distinct trends in polar Octadecabacters seem to be an increased incorporation of negatively charged amino acids (both D and E), as well as valine (V) into helices (Figure 3.24). At the same time, their predicted helices contain a lower percentage of isoleucine, methionine, asparagine (N), and lysine.

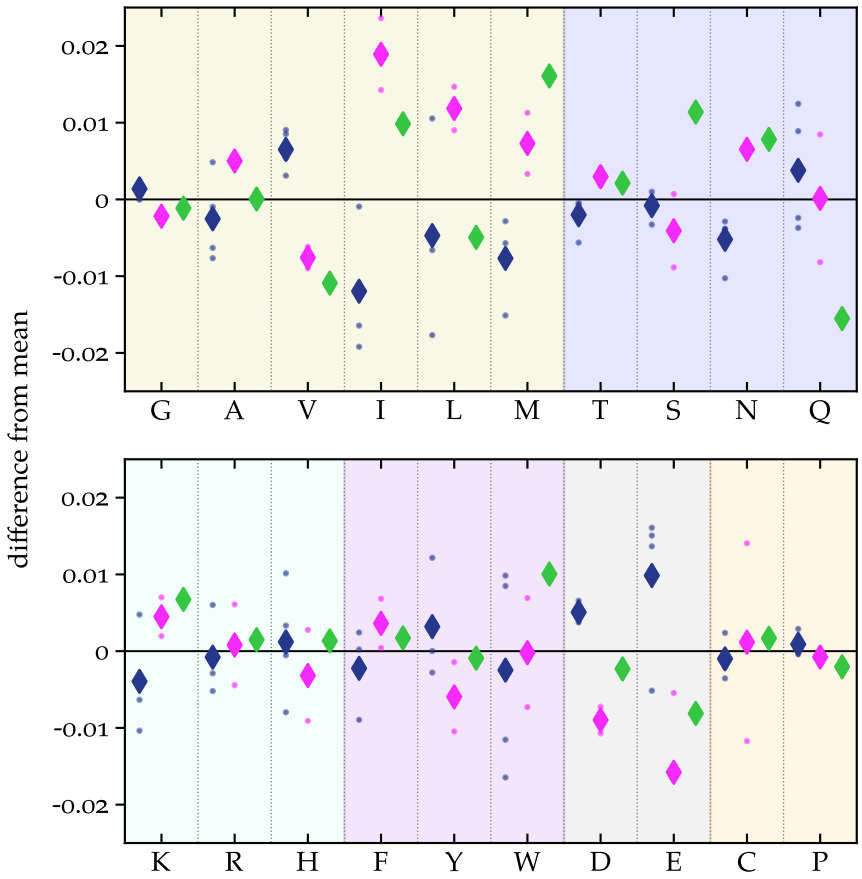


Figure 3.24: Differences in α-helix amino acid content of Octadecabacter groups. The average number of each amino acid in core-protein positions predicted to be part of a helix was related to its average number outside predicted helices for each group. The plot shows the difference between these mean ratios and the base (mean) ratio over all seven Octadecabacter strains with similar GC-content (Table 3.1). Octadecabacter groups and colours as defined in Section 3.2 (I, II, and IV); amino acid groups are the same as in Figure 3.23.

3.9 OCCURRENCE OF OCTADECABACTER-RELATED SEQUENCES IN METAGENOMES

A total of 591 metagenomes on the mg-RAST platform met the inclusion criterion for the analysis (Section 2.12.1), of which 559 could be successfully downloaded and mapped. The metagenomes contained either reads, or assembled contigs, with a total of 1,342,811,866 sequences assigned to the bacterial kingdom by mg-RAST. Mapping returned 359,928 sequences which aligned to the Octadecabacter genomes. The filtering procedure focusing on distinction between Octadecabacter groups (Section 2.12.3) excluded 20,388 hits (≈5.6 %), whereas 23,983 sequences (≈6.7 %) were excluded when pairwise comparisons between all Octadecabacters were considered. All relevant data are also provided in the supplementary files (see Appendix A.1).

For 522 metagenomes, geographic coordinates were available. They stemmed from 214 distinct sampling points, at 113 of which several dates, size fractions, depths, etc. were sampled, leading to 308 metagenomes with duplicated coordinates. Eighty-two of the stations contained metagenomes with sequences mapping to at least one Octadecabacter, while 132 stations returned no hits (Figure 3.25).

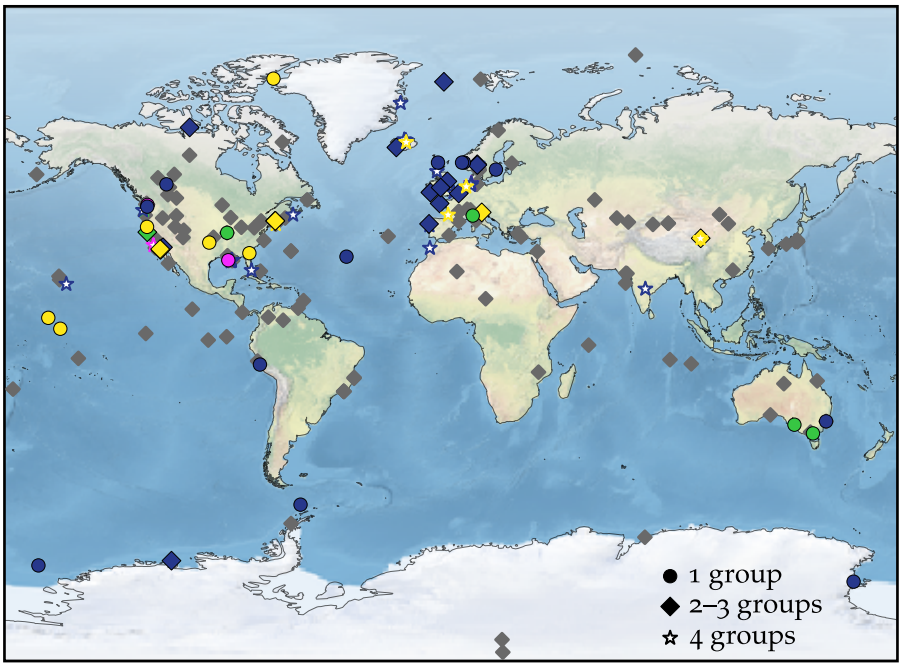
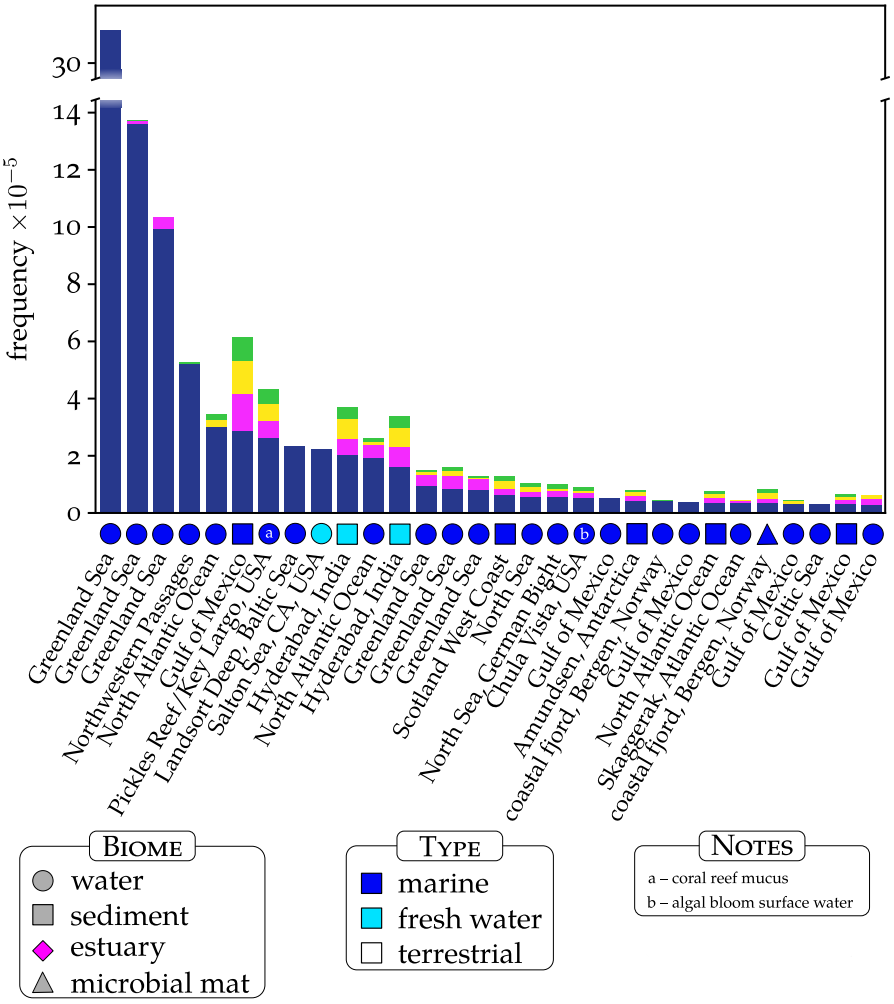


Figure 3.25: Octadecabacter groups at metagenome sampling sites. Where multiple metagenomes were present for a single sampling point, they were merged using the highest relative abundance of each group in any of the samples. Shown are sampling sites with reads mapping to no Octadecabacter (132 stations, grey diamonds), one specific Octadecabacter group (24 stations total, circles of correspondent colour), 2–3 Octadecabacter groups (diamonds, coloured according to most abundant group) or to all Octadecabacter groups (37 stations, stars coloured according to most abundant group).

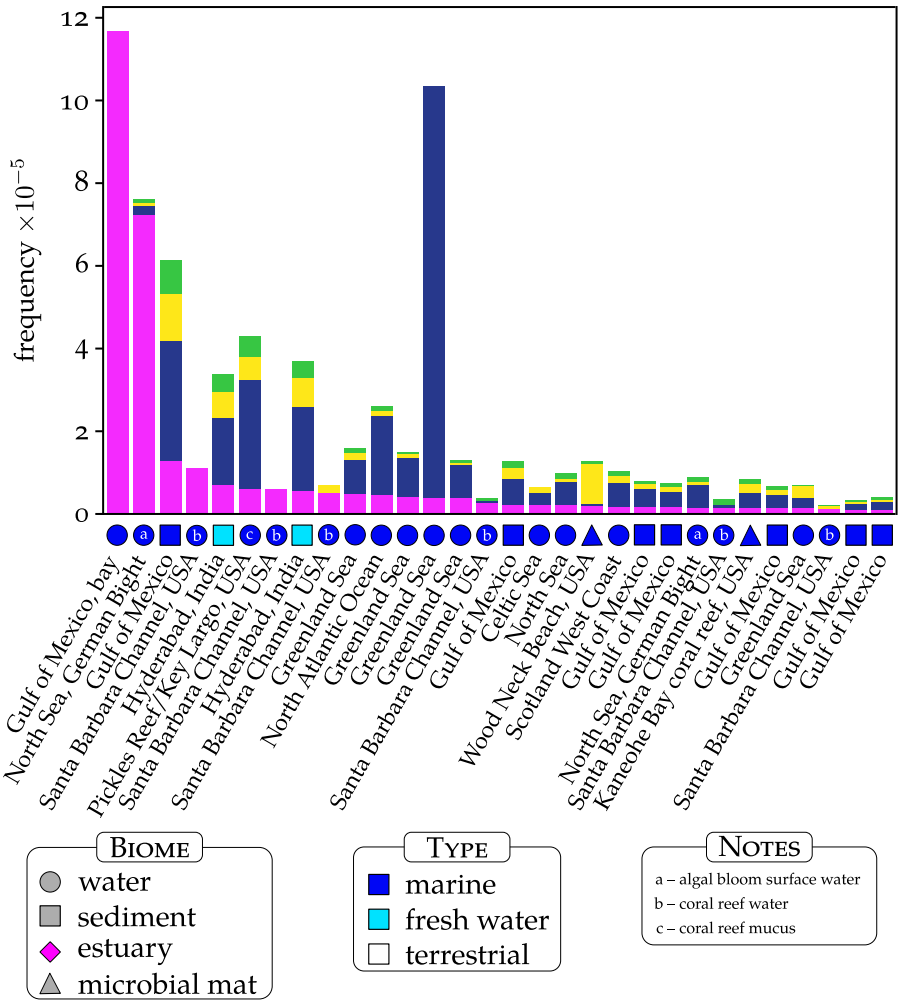
Polar Octadecabacters (group I) were detected in 117 metagenomes at 69 sampling sites and are the most widely distributed of the four groups (Figure 3.25). Likewise, they are often the most abundant of all detected Octadecabacter groups (Figure 3.26). Groups II, III, and IV were detected in 74, 86, and 88 metagenomes at 51, 49, and 60 stations, respectively. At 37 sampling sites, metagenome sequences mapped to all four Octadecabacter groups, while 24 stations featured one group exclusively (Figure 3.25).



(a) Octadecabacter group I

Figure 3.26: Metagenomes ranked by abundance of Octadecabacter groups. The number of nucleotides in reads mapping to a specific Octadecabacter group, divided by the total number of nucleotides assigned to the bacterial kingdom, is plotted as the frequency. Sub-figures (a) to (d) list the 30 metagenomes, in which groups I to IV occur with the highest frequency, respectively, along with information on the biome, sample type, and sampling location.

The vast majority of metagenomes with a high frequency of nucleotides mapped to groups I and II originates from marine water samples, with marine sediment being the second most frequent biome

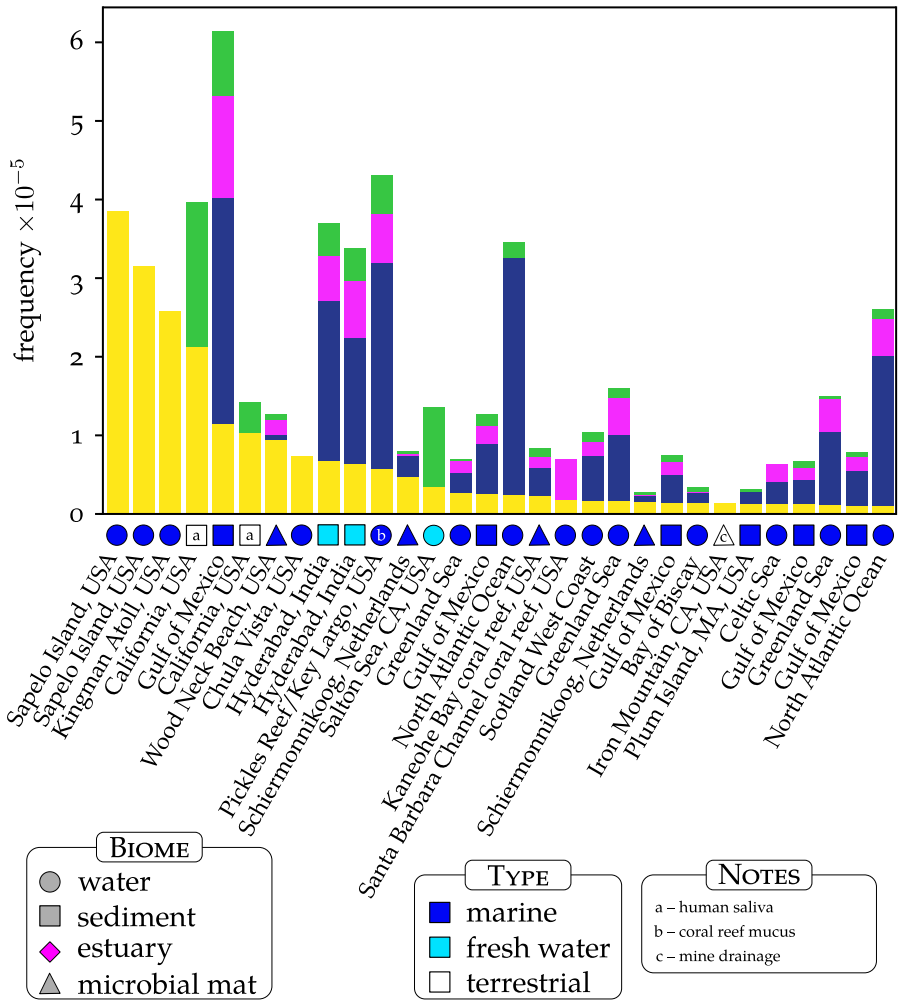


(b) Octadecabacter group II

Figure 3.26: Metagenomes ranked by abundance of Octadecabacter groups (continued)

(Figures 3.26a and 3.26b). Conversely, habitat types are more evenly distributed among the metagenomes with nucleotides mapping to groups III and IV, showing a higher number of microbial mats, estuaries, or terrestrial samples (Figures 3.26c and 3.26d). These groups' frequencies are, however, generally much lower. Groups I and II, on the other hand, predominantly feature in 4 and 2 of the 6 metagenomes with the highest frequency of nucleotides mapped to Octadecabacters, respectively.

While most metagenomes listed in Figure 3.26 contain reads mapped to multiple, or all Octadecabacter groups, those which show the highest frequency of an individual group are usually exclusive to that group. Therefore, despite the frequent co-occurrences, it is possible to identify some habitat preferences of the different groups. For example, group I most prominently features in water samples of the Greenland Sea and the North Atlantic Ocean in general, i. e. water samples of high



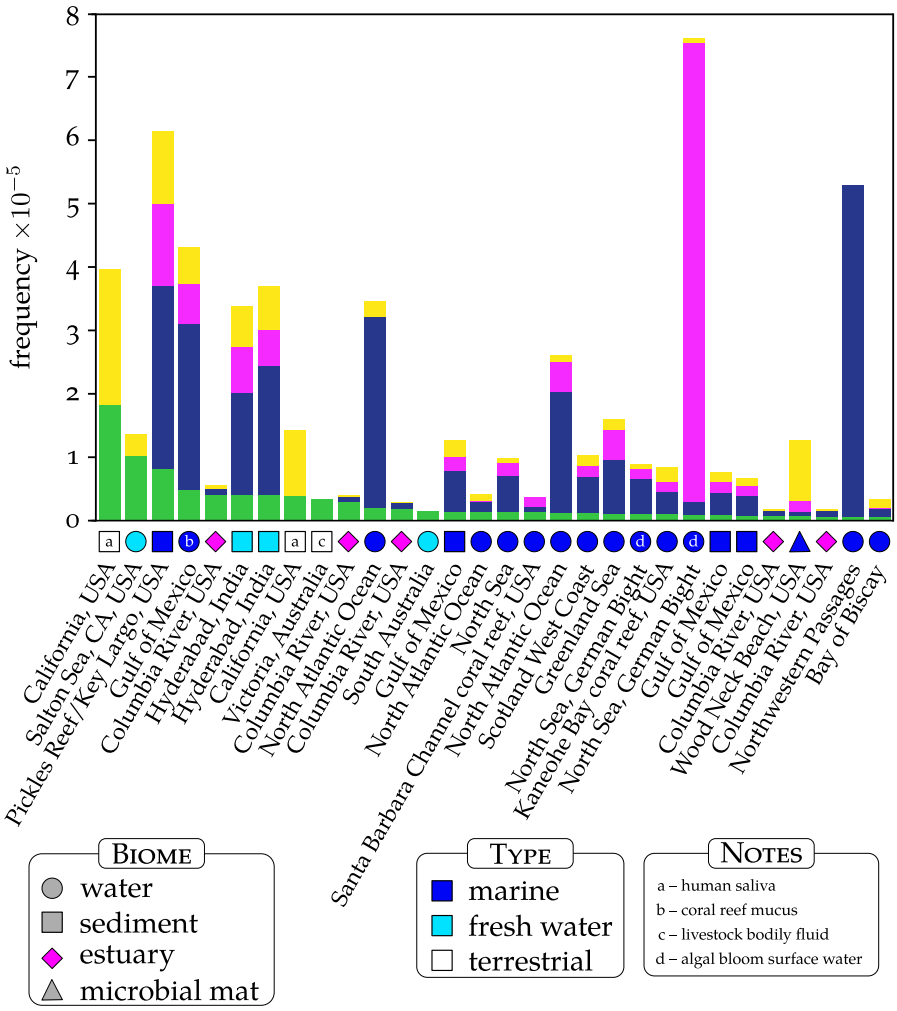
(c) Octadecabacter group III

Figure 3.26: Metagenomes ranked by abundance of Octadecabacter groups (continued)

latitudes, where other Octadecabacter groups seem to be nearly absent (Figure 3.26a).

Group II has more hits in metagenomes from coastal areas and bays, and notably features in water samples of coral reefs or algal blooms, where it likewise is often the only detected group (Figure 3.26b). These regions are mostly positioned at intermediate latitudes, but the group also seems to occur to some extent at higher latitudes, e. g. in water samples of the North Atlantic Ocean.

Group III shows the highest frequencies of mapped nucleotides in several water samples from low latitudes (Figures 3.25 and 3.26c). Nonetheless, there are also highly similar sequences present in metagenomes from higher latitudes. Interestingly, it is the only group detected in one of the northern-most samples included in this analysis, a freshwater lake microbial mat on Ellesmere Island, Canada (Figure 3.25).

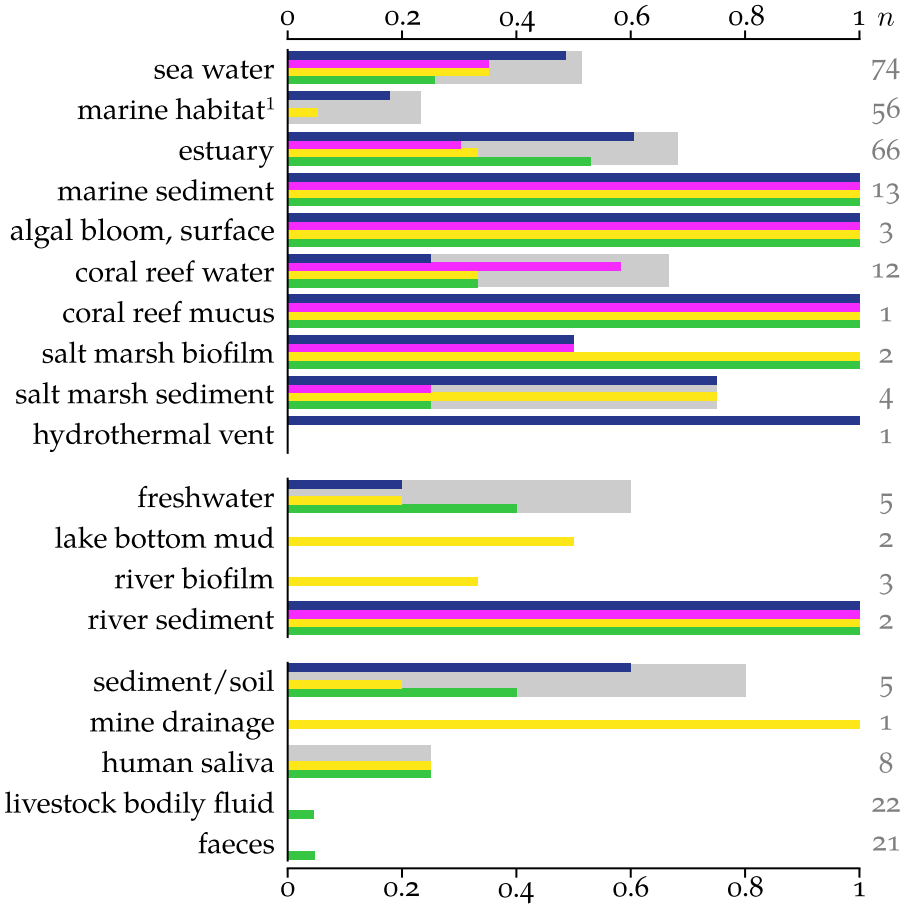


(d) Octadecabacter group IV

Figure 3.26: Metagenomes ranked by abundance of Octadecabacter groups (continued)

Group IV generally exhibits the lowest number of mapped reads within the Octadecabacters, and among the metagenomes with the highest nucleotide frequency of this group, there seems to exist no particular habitat preference (Figure 3.26d). Notably, it is featured more prominently in some of the tested estuarine metagenomes than the other groups.

In general, when evaluating how many of the candidate metagenomes actually contained reads mapping to Octadecabacters, some differences between the habitat types are noticeable. The sea water and "marine habitat" biome categories, for the latter of which no further details were provided in the metadata, contained among the most metagenomes, based on the biome and sample type descriptions. However, only 25 to 50 % of these gave hits in the stricter mapping procedure (Figure 3.27). On the other hand, all samples of marine and freshwater



[1] – sample metadata contained no further details

Figure 3.27: Octadecabacter group occurrence in different biomes. For each biome/sample type represented by a metagenome in this analysis, the proportion of metagenomes in which individual Octadecabacter groups were detected is plotted (colours as in Section 3.2, groups I, II, III and IV). Grey numbers (n) give the total amount of metagenomes for each biome/sample type. Where more than one group occurs in a habitat, a grey bar indicates the fraction of metagenomes with reads mapping to any group. Biomes are sorted into marine (top), freshwater (middle), and terrestrial (bottom) habitats.

sediment, in which mg-RAST assigned sequences to the *Octadecabacter* genus, also returned mapped reads.

Agreement in predictions of Octadecabacter occurrence between mg-RAST and the mapping approach used here is generally high for marine biomes (70 to 100 %, with the exception of sea water and "marine habitat"). It is, however, lower in freshwater and terrestrial biomes, where mg-RAST predicted Octadecabacter-related sequences in 12 and 57 metagenomes, respectively (including bodily fluids and faeces, Figure 3.27). Read mapping returned hits in 7 and 9 of these metagenomes, respectively. In over half of the sample types, reads mapped only to either group III, or group IV. Sediment (both freshwater and terrestrial), and fresh water samples pose a notable exception

in this regard, having a higher proportion of metagenomes with reads mapping to multiple Octadecabacter groups (Figure 3.27).

4 DISCUSSION

4.1 OCTADECABACTER LIFESTYLE AND GENE CONTENT DIVERGENCE

Chapter 3 presented several genomic characteristics of Octadecabacter clade bacteria and made initial comparisons to other members of the *Roseobacter* group. These data enable an investigation, in how far such characteristics generally relate to differences in lifestyle between the analysed strains, and whether Octadecabacters constitute special cases among the roseobacters.

Having reconstructed the group's most likely phylogeny using a phylogenomics approach is particularly useful in this regard, as it allows comparison of similarities in gene content to phylogenetic relation in potentially large groups of genomes. Section 4.1.1 will evaluate the obtained phylogenies reliability and discuss their congruence with observed lifestyles and genetic divergence. The discussion will cover both the *Roseobacter* group as a whole, and the Octadecabacters in particular.

Variability in gene content is generally well captured by analysing a group's pan-genome. Such an analysis provides standardised characteristics, which can be readily compared between different organism groups. In Section 4.1.2, results of the Octadecabacters' pan-genome analysis will be put in relation to other clades of the *Roseobacter* group.

Finally, genus delineation is a widely debated topic, and can be particularly problematic for the functionally diverse roseobacters (see Sections 1.4 and 1.5). Section 4.1.3 will address this problem, using the example of the Octadecabacter clade to evaluate and compare several metrics proposed for use in genus delineation.

4.1.1 *Phylogeny*

As stated in Section 3.2, a super-matrix-based MLSA of the *Roseobacter* group core-genome generally reproduced the topology observed in other studies [59, 60]. The Octadecabacters form a distinct and highly supported clade (compare Figure 3.2, p. 30), with isolate NH9-P7 being the earliest branching and therefore most distant member, followed by *P. jejudonensis*. The polar isolates cluster together with high bootstrap support, and also form a common sub-clade with the temperate members *O. temperatus* and *O. ascidiaceicola*. However, phylogenies derived from super-matrices are under criticism for not necessarily reflecting the most prevalent topology, as well as rendering bootstrap supports misleadingly high [286].

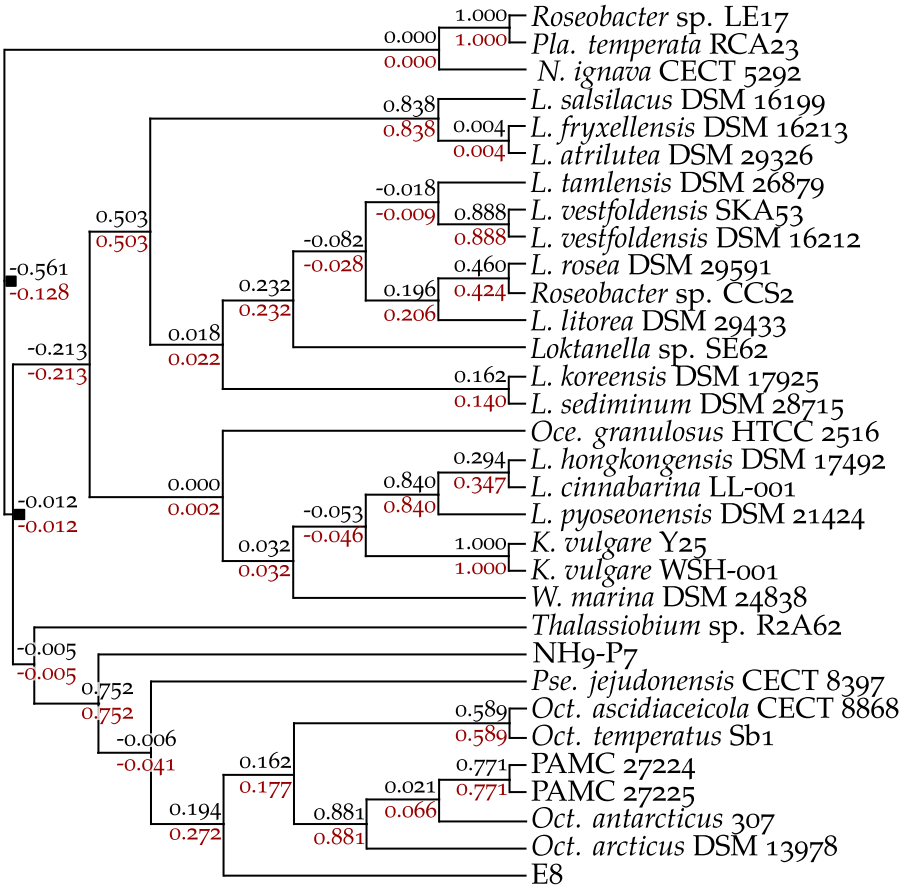


Figure 4.1: Reassessment of bipartition support within the super-matrix-based tree topology, using IC and TC values. The tree topology was taken from Figure 3.2, and support values at each bipartition were re-estimated as IC (black) and TC (red) by comparing the MLSA with 139 single core-gene phylogenies (see Section 2.6 for details).

A more realistic estimate of bipartition support can be derived from the frequency in which the bipartitions occur in the ML phylogenies of the individual genes. More precisely, the frequency of each bipartition can be related to the two most frequent conflicting bipartitions, which results in the information theory-based internode certainty (IC) and tree certainty (TC) values [174]. Values close to 0 indicate that the next two most frequent topologies are almost as prevalent as the one shown in the tree, whereas values close to 1 show the absence of such conflicts (negative values mean that at the respective site, other bipartitions occur more frequently altogether).

When comparing the 139 core-gene phylogenies with the super-matrix-based one in this manner, it is obvious that the high MLSA bootstrap support values are exaggerated, implying a false sense of certainty in the final tree topology (Figure 4.1). While the formation of an Octadecabacter clade, distinct from their closest relative, *Thalassiobium*, is still well supported (IC value of 0.75), the branching order between the more distant members *P. jejudonensis*, isolate NH9-P7,

and isolate E8, appears to be much less reliable than indicated by the bootstrap analysis (signified by low IC values of 0.006 to 0.2 at the respective bipartitions). Common ancestry of the four polar isolates, *O. temperatus*, and *O. ascidiaceicola* is seemingly conflicted in the individual *Roseobacter* group core-gene trees (low IC of 0.16), which agrees with the network representation of 1,513 Octadecabacter clade core-gene phylogenies (Figure 3.3). At the same time, both the existence of a polar clade, and the grouping of *O. temperatus* and *O. ascidiaceicola*, maintain a high support.

Over the rest of the *Roseobacter* group topology, certainty is comparatively low: multiple negative IC values indicate that the MLSA phylogeny does in fact not represent the phylogeny of most of the component genes (see Figure 4.1).

It should be noted that the genes used for MLSA were selected solely based on their classification as single-copy core-genes, and not individually tested for the strength of their respective phylogenetic signal. Consequently, part of the observed uncertainty may simply result from phylogenetic noise within the individual gene trees. In addition, a super-matrix-based phylogeny always constitutes a compromise between conflicting phylogenetic signals of its component genes, and is therefore inherently uncertain by necessity. This becomes clear when analysing the Robinson-Foulds (RF) distances between the super-matrix-based MLSA phylogeny and the individual gene trees, and comparing them to the pairwise RF distances of the individual trees among each-other. In this case, the median normalised RF distance to the MLSA phylogeny is lower (by 0.13) than the median normalised distance of pairwise comparisons of the individual gene trees (see Figure 4.2). Thus, the MLSA phylogeny appears to present an adequate compromise between the conflicting single gene trees, minimising the overall distance, while not necessarily agreeing with the majority of their topologies. For this reason, it is appropriate to use this phylogeny as the basis of ancestral state reconstruction and the phylogenetic birth-and-death model, as described in Section 2.9, although the huge uncertainties concerning the bipartitions (especially deeper ones) should be kept in mind when discussing findings which rely on phylogenetic data.

Roseobacter group members are widely distributed, mostly in marine environments, where they occupy a variety of ecological niches [62]. Even within the genera of this group, different species are often found to dwell in different micro-environments, which may be as diverse as planktonic, biofilm, or host-associated habitats (compare Table A.1, p. 125). In fact, some argue that this colonisation of new micro-environments, and the subsequent decrease of gene transfer between populations, could constitute a mechanism of speciation [287]. The high genetic flexibility frequently observed in *Roseobacter* group members likely makes such a transition easier, and thus contributes to this group's adaptability [69].

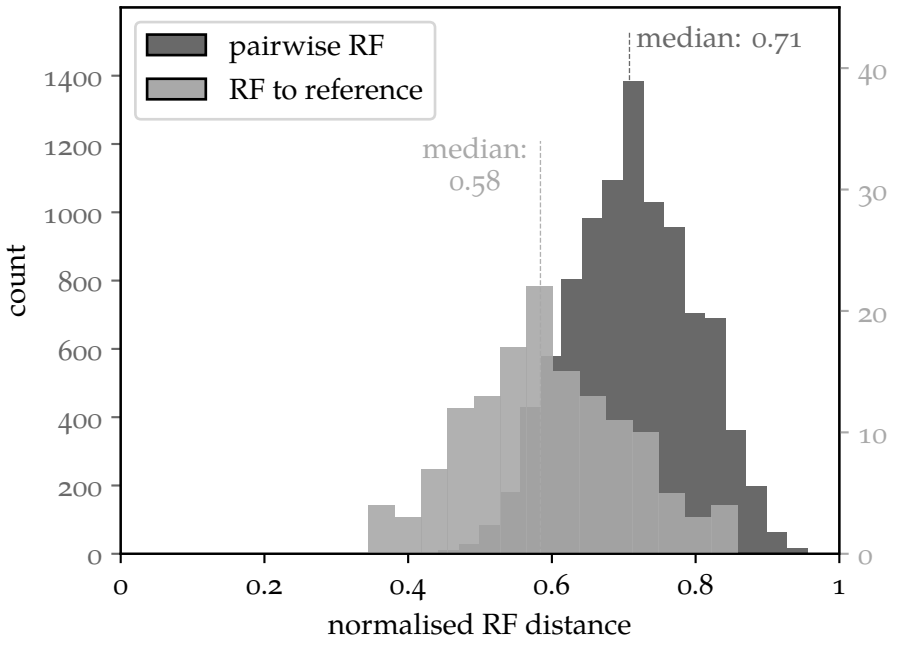


Figure 4.2: RF distances between individual core-gene and super-matrix-based phylogenies in the *Roseobacter* group. Pairwise RF distances were calculated for all 139 individual core-gene phylogenies, and their distribution is shown in dark grey (left axis). The distribution of their distances to the super-matrix-derived phylogeny is depicted in light grey (right axis).

Despite such varying lifestyles, gene content-based hierarchical clustering showed high consistency with the MLSA phylogeny (see Figure 4.3). Clades defined in ref. [59] are largely conserved between both trees, with only few individual organisms grouping differently in the gene content-based clustering. Most differences in bifurcations between the two trees are restricted to deep nodes, at which point the aforementioned uncertainty in the MLSA branching forbids exact comparisons. Consequently, vertical inheritance appears to dominate gene content over horizontal acquisition, i. e. closely related *Roseobacter* group members possess a distinct genetic coherence. Notably, this coherence does not seem to be limited to the core-genome. The removal of core-genes within clades throughout the *Roseobacter* group and subsequent hierarchical clustering results in a tree, which is even more congruent with the MLSA phylogeny (Figure A.3, p. 133; normalised RF distance of 0.47 without sub-clade core-genes, versus 0.59 when including core-genes). This is even the case in groups like the polar Octadecabacters, which are known to have highly flexible genomes with strong potential for lateral gene transfer (LGT).

The cloud-, and shell-genomes reflect an organism's adaptations to its particular micro-habitat [288]. As these adaptations occur predominately via acquisition of laterally transferred genes [289], these acquired genes will primarily make up the shell-genome. Since most roseobacters possess genes encoding for GTAs (Section 1.3), within-

lineage transfer of genetic material may have a role in shaping individual strains' gene contents [60]. Nonetheless, similarities in cloud- and shell-genome content do not reflect similar habitats, or geographic proximity[1] of the compared *Roseobacter* group strains (Table A.1), but, as stated above, rather follow the phylogeny (Figure A.3). Despite the higher number of genomes compared here than were available to Newton et al. [60], reliably testing the idea of a common roseobacter gene pool, the exchange of which facilitates environmental adaptation, would likely still require more sequenced representatives [60].

In the phylogenetic analysis based on the Octadecabacter core-genes, a network representation was chosen to directly visualise conflicts in phylogenetic signal between the individual gene trees (Figure 3.3). With regard to the network's overall fidelity, an analysis of RF distances, equivalent to the one performed for the *Roseobacter* group MLSA, showed a lower median distance of the individual gene phylogenies to the network than to each-other (0.33 vs. 0.50, respectively), indicating that the network may be an accurate representation of the phylogenetic relationships within the Octadecabacters.

Clustering the nine strains based on their gene content, and thereby their potential functional adaptations, revealed some notable differences (Figure 3.3). While the polar clade and the clade formed by *O. temperatus* and *O. ascidiaceicola* remain individually unchanged by gene content clustering , they no longer form a common larger cluster as in the core-genome phylogeny (albeit at low IC support). Within the polar species, *O. antarcticus* clusters with *O. arcticus*, rather than with the other Antarctic isolates. This may reflect their similar lifestyle: *O. arcticus* and *O. antarcticus* occur in sea water as well as in sea ice, while the PAMC isolates were procured from marine sediments (see Section 1.4). Notably, this particular bifurcation is also conflicted in the network, meaning that a significant proportion of single gene phylogenies puts *O. antarcticus* in closer relation to *O. arcticus* than to the PAMC isolates.

Vollmers et al. attributed the high observed similarity between both strains to a mutual genetic exchange, postulating a potential connection between both Arctic and Antarctic populations [77]. Apart from the coherent grouping of the polar Octadecabacter clade members, lifestyle does not seem to be a good predictor of content-based similarity, neither does geographic proximity: *O. temperatus* and *O. ascidiaceicola* are stably grouped together, despite the fact that they were isolated from distant parts of the world, and from different marine micro-environments (free living fraction of the water column for the former, sea squirt-associated for the latter, see Section 1.4). Likewise, both the isolates from Korea (*O. ascidiaceicola* and *P. jejudonensis*) and

1 hierarchical clustering based on the geographic distance of the available isolation coordinates (Figure A.4) returned a tree with an RF distance of 0.96 to the MLSA, as well as both the gene content and core-genome-reduced gene content trees

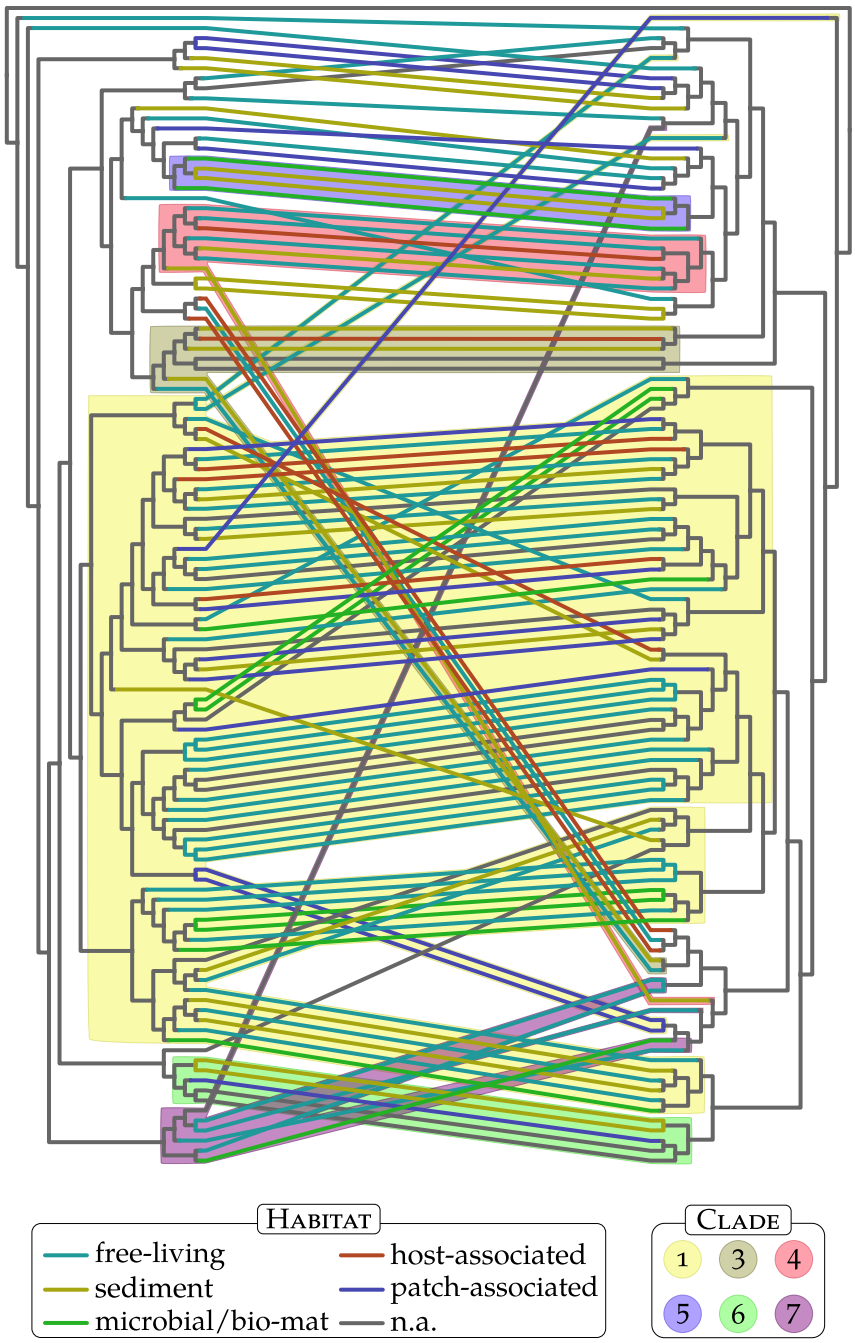


Figure 4.3: MLSA phylogeny of 115 roseobacter genomes plus outgroup on the left, as given in Figure A.2, versus hierarchical clustering of the same genomes on the right, based on pairwise Jaccard distances of OG presence and absence. Clades defined in ref. [59] are coloured as in Figure A.2. Each strain's lifestyle or preferred type of habitat was estimated from the source publication (see Table A.1), and branches are coloured accordingly. RF distance between both trees is 0.59, determined using `ete3` [177].

those from the North Sea (*O. temperatus* and isolates NH9-P7 and E8) do not form common clusters.

This reflects the above observations for the whole *Roseobacter* group, namely that phylogeny is a better predictor of content-based similarity than similar lifestyles. In the present case, isolate NH9-P7 and *P. jejudonensis* are both phylogenetically distant (Figure 3.2), as well as highly dissimilar from the other clade members on the sequence level (Figure 3.1). Due to this distance, vertical inheritance of genes may determine their position in content-based clustering, rather than similarities in lifestyle. As with the whole *Roseobacter* group, a higher number of available genome sequences would allow for more conclusive statements, how far geographic proximity and/or similarity in lifestyle correlate with phylogenetic relation. It should also be noted that the circumstances under which an organism is isolated do not necessarily reflect its preferred lifestyle, and this information is therefore associated with some uncertainty.

In general, the observed distances in gene content between the Octadecabacters are relatively high, which may be partly owed to the fact that singletons were included in the distance calculation. The polar Octadecabacters are, as a group, most divergent in this regard. Apart from them, strain E8 is notably the single most distantly branching isolate of the Octadecabacter clade (based on its distance to the tree root). This is also reflected in both groups' singleton fractions (Figure 3.6), as will be discussed in the following section.

4.1.2 *Pan-genome analysis*

The genomic diversity of a group of organisms can be assessed by analysing the correlation between the total number of different orthologous groups (OGs) (the so-called pan-genome) and the number of compared genomes [290]. This correlation usually follows a power law. Pan-genome analysis of the Octadecabacters revealed that the corresponding fitted power law curve does not reach saturation (exponent $0 < b < 1$, see Table 3.3, Section 3.3.1), a property which defines an open pan-genome [290]. This means that the currently available Octadecabacter clade genome sequences do not yet fully represent the clade's inherent genetic diversity [188]. Instead, when added to the analyses consecutively, each new genome contributes ≈700 novel genes to the clade's pan-genome (Figure 4.4), reflecting unique adaptations to their different habitats (compare Section 1.4), and indicating high flexibility of the accessory genome for acclimatisation to specific niches [291]. This adaptability is, however, not more pronounced in the Octadecabacters than in other clades within the *Roseobacter* group. The analysed comparison clades show similar power law exponents (see Table 3.3), and the number of novel OGs added to the pan-genome by the successive addition of each new species does not differ significantly

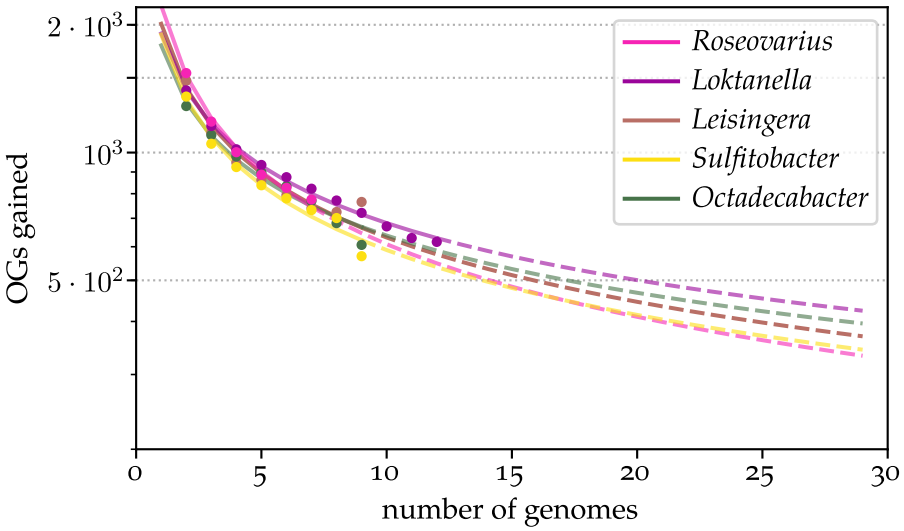


Figure 4.4: Change in pan-genome size in relation to the number of comparison genomes for Octadecabacters and other comparison clades of the *Roseobacter* group. Clades were selected based on similar phylogenetic branching depth (see Table 3.2). Median values fitted to a power law as in Equation (2.1), p. 23. Dashed lines are extrapolations based on the parameter estimates.

between any of the investigated clades (compare extrapolated curves in Figure 4.4). Consequently, they possess equally broad pan-genomes, which supports the frequently stated assumption that high genome plasticity and genetic adaptability are general traits of the *Roseobacter* group [69, 93] (Section 1.3).

The *Roseobacter* group possesses an open pan-genome, which comprises >70,000 OGs in 114 genomes (Figure 3.5 and Table 3.3). Generally, open pan-genomes are common for bacteria at the genus [292–294], and the species level [188]. The increase in number of genes by successively added genomes is a relatively simple metric that can easily be compared between different studies. Interestingly, this metric is notably higher in the comparison clades of the *Roseobacter* group, including the Octadecabacters, than for other analysed bacterial groups [188, 292–294]. One should keep in mind that observable pan-genome trends are dependent on the number and nature of the compared genomes [188]. Due to the limited amount of data for the roseobacters, and Octadecabacters in particular, the observed curves can only be taken as a trend. While the estimated curve parameters seem robust due to their low variances (Table 3.3), they do not reflect the inherent uncertainty introduced by the selection of comparison genomes. In order to more conclusively demonstrate that the Octadecabacters possess a higher genomic diversity than other comparison groups, the analyses would therefore need to be repeated with more data. In addition, an open pan-genome on the genus level can mask the closed pan-genomes of one or multiple constituent species [293]. Given more genomic data, it would be interesting to see whether lineages exist within the roseobac-

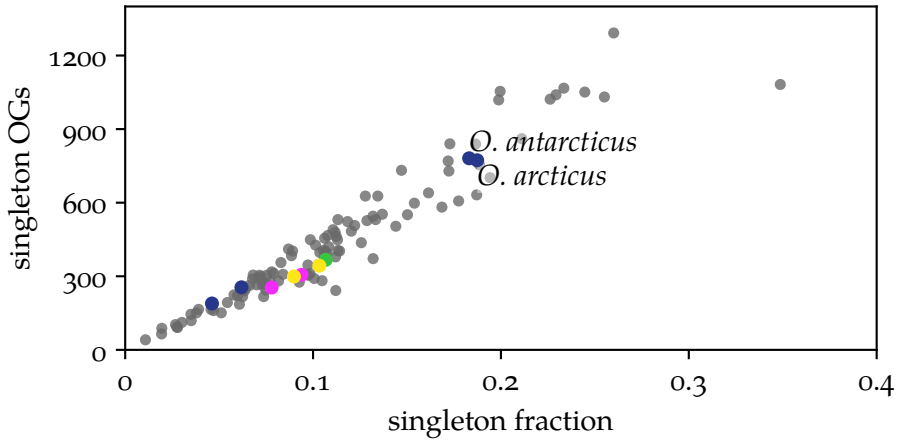


Figure 4.5: Occurrence of singletons in *Roseobacter* group genomes. The number of singletons (i. e. orthologous groups (OGs) exclusive to one genome) is plotted against their fraction of total OGs for each comparison genome listed in Table 2.2, excluding the outgroup genome *Parvularcula bermudensis* and *Phaeobacter* sp. LSS9. Octadecabacters are highlighted in colours corresponding to their groups defined in Section 3.2.

ters, which do not fit into the *marine generalist* scheme, and if so, at which level of phylogenetic divergence this behaviour would emerge.

Apart from allowing an estimation of general genome diversity, gene orthology information enables a detailed look at the specific distribution of genes within a selected group of organisms. To that end, the distribution of genes in the core, cloud, and shell-genomes of the Octadecabacters, as well as several comparison clades, was analysed (Section 3.3.1). As noted there, *O. arcticus* possesses the largest shell-genome fraction among the Octadecabacters (Figure 3.6). An earlier study found both *O. arcticus* and *O. antarcticus* to contain the highest number of unique genes among 32 *Roseobacter* group genomes [60]. Among the 115 comparison genomes used here, they do not represent extreme cases in this regard, as multiple other strains possess more singletons, as well as a higher fraction of unique OGs in their genome (Figure 4.5).

singleton – gene exclusive to one genome of a comparison group

Nonetheless, the core-genome still represents the largest genome fraction in all Octadecabacters (40 % in the polar Octadecabacters and 50 % in the rest, Figure 4.6). The shell-genome of *O. arcticus* makes up ≈30 % of its CDSs. Other clades within the *Roseobacter* group show similar variation in the three categories (Figure 4.6). For most of the respective genomes, the core-genome comprises 30 to 50 % of the total number of CDSs (varying according to genome size). Although the Octadecabacters tend towards relatively smaller cloud-, and larger shell-genomes than their relatives, they do not constitute extremes in any of the two categories. For example, some Roseovarius clade members contain even smaller cloud-genomes than the Octadecabacters, and the shell-genome fraction of one Loktanella clade strain exceeds that of *O.*

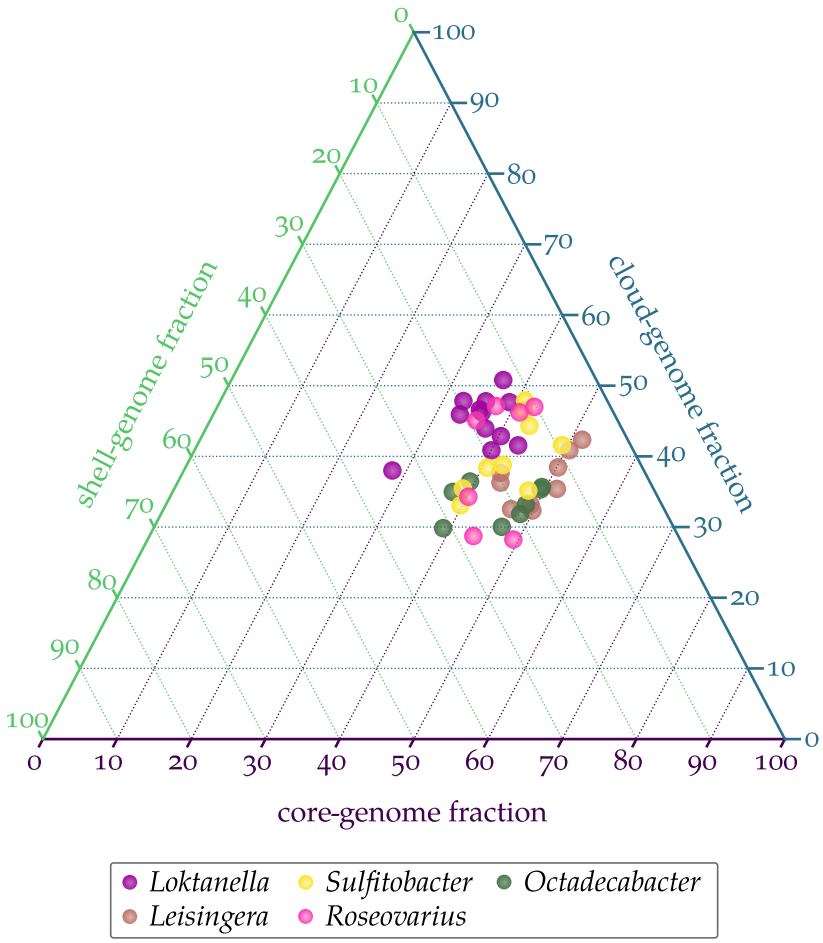


Figure 4.6: Core-, cloud-, and shell-genome fractions (in percent) among selected *Roseobacter* group lineages. Comparison lineages were selected based on similar branching depths compared to the Octadecabacter clade (Table 3.2). In the case of extremely similar genomes (i. e. near-zero branch lengths in the reference phylogeny), such as the two PAMC isolates, only one respective representative was included.

arcticus. The exact genome category distribution is markedly different between the Octadecabacters and the closely related neighbouring Loktanella clade: *Loktanella* strains tend towards larger cloud-genomes, at approximately equal expense of both the shell-, and core-genomes.

4.1.3 *Genus delineation within the Octadecabacter clade*

The extensive functional divergence discussed in the previous sections complicates a sensible delineation of genera within the Octadecabacter clade and *Roseobacter* group in general. The re-classification of *O. jejudonensis* as *Pseudooctadecabacter* was, for example, mainly derived from its phenotypic differences to three other clade members in the polyphasic analyses [79]. Based on the genomic flexibility and range of habitats, however, such phenotypic differences can accumulate quickly, and

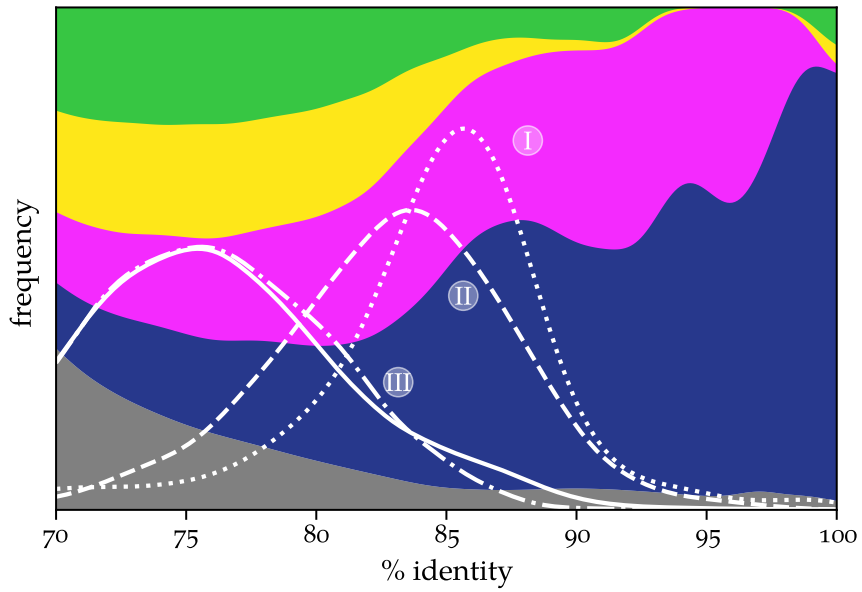


Figure 4.7: Nucleotide identity distribution within Octadecabacter groups. White curves show distribution of nucleotide numbers over %-identity values for pairwise alignments of 1,000 nt-fragments within Octadecabacter groups I to III (labelled accordingly), and between all nine genomes (solid line). Background visualises how Octadecabacters compare to other roseobacters (Table 2.2), showing what proportion of nucleotides at a particular %-identity value to any Octadecabacter stem from a roseobacter (grey), or another Octadecabacter (coloured according to its group). Note that rRNA genes were excluded from the comparison (Section 2.12).

traits such as carbon compound utilisation, among others, show a high phylogenetic dispersion [295]. Digital DNA-DNA hybridisation (DDH) demonstrated a low similarity between *P. jejudonensis*, isolates NH9-P7 and E8, and the other clade members (Figure 3.1), and on this basis it seems justifiable to divide this clade into at least two genera.

Similarly, the Octadecabacters' ANI profile is distributed around a lower mean than that of the intra-genus range observed for other organisms [85] (Figure 4.7). This metric would suggest division into five distinct genera. Two of these are formed by the members of groups I and II, as between them, the nucleotide identity within their compared fragments peaks around 85 % (Figure 4.7, compare to data in ref. [85]).

However, the suitability of sequence-based measures for genus delineation, particularly that of ANI, has recently been questioned by some authors (e. g. ref. [86]). Instead, Qin et al. propose the percentage of conserved proteins (POCP) as an alternative, defining a genus as the "group of species with all pairwise POCP values higher than 50 %" [86]. They consider two proteins as conserved when these show reciprocal BLAST hits above a defined score, identity, and alignment length. The orthologue detection performed here used similar criteria (minimum 35 % identity, versus 40 % in Qin et al., and 65 % alignment coverage,

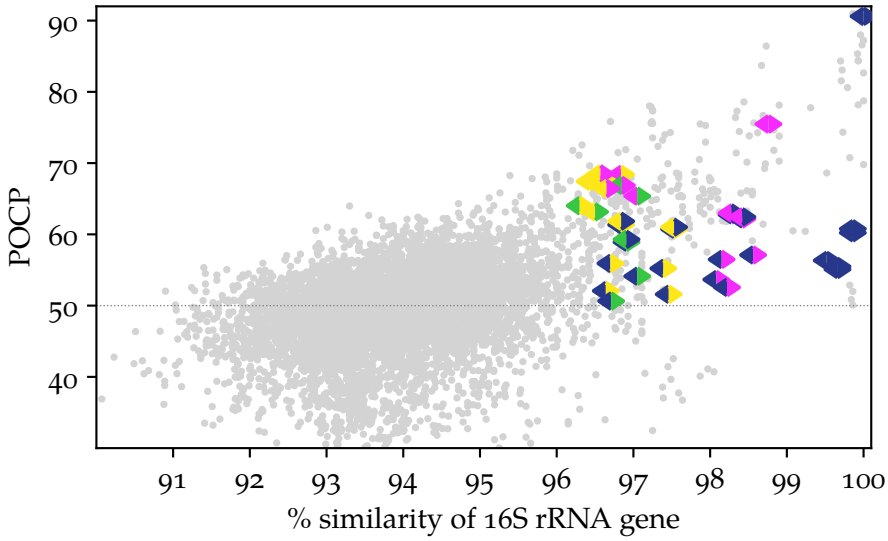


Figure 4.8: Percentage of conserved proteins (POCP) versus 16S rRNA gene similarity in roseobacter strains. POCP in this case refers to the number of shared OGs, divided by the total number of genes in each two compared genomes. Data for Octadecabacters are represented with larger symbols. Symbol colours indicate the groups (Section 3.2), to which the two compared Octadecabacter strains belong. One proposed definition suggested that a genus comprise the set of bacterial strains in which all pairwise POCP are higher than 50 % [86].

versus 50 % [86]), so that comparison of the inferred orthology relationships should reflect the POCP well. When applying the proposed genus definition to the current case, it becomes clear that it is not suitable for the Octadecabacters, or the *Roseobacter* group in general, at least not when using the stated criteria for determining conserved proteins.

The core-genome constitutes the largest fraction of all Octadecabacter strains' CDSs (see previous section), and all nine consequently show pairwise POCPs of over 50 % (Figure 4.8). However, assignment of at least *P. jejudonensis* and isolate NH9-P7 to a different genus than the rest is strongly supported by multiple factors, including their distant clustering in the phylogeny, the low sequence similarity (Figure 3.1), and their variation in GC-content (Table 3.1), in addition to the differences observed by Billerbeck et al. The generally high POCPs reflect the coherence in genetic equipment discussed in Section 4.1.1. Notably, when compared among each other, groups II to IV show POCP values well over 60 % (Figure 4.8). This again underlines the polar group's divergence concerning gene content. It also demonstrates how vulnerable an individual metric can be to the boundary conditions presented by a particular group of organisms.

Ultimately, rather than basing it on one set standard, the definition and delineation of genera should be based on multiple criteria and should follow a careful individual assessment, taking into account the specific properties of the group of organisms in question. For example,

keeping in mind that functional features like carbon compound utilisation, nitrogen fixation, or even particle colonisation [287, 295] are phylogenetically dispersed, their distributions among bacterial strains are unsuited to define genus or species boundaries in functionally versatile organism groups with high genome flexibility.

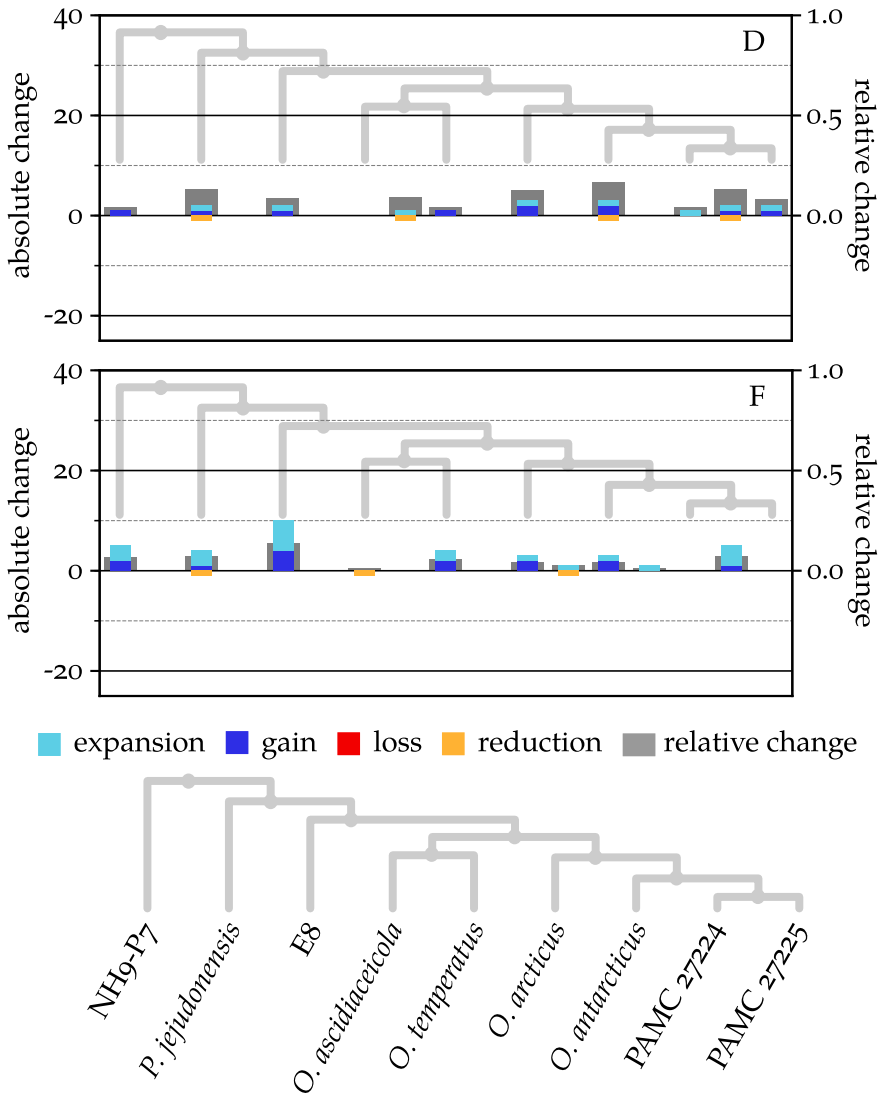
4.2 FUNCTIONAL GENE CONTENT AND ITS EVOLUTION

Polar Octadecabacters show some notable differences in their functional gene content to the temperate isolates (as captured by the COG classification, Figure 3.7 p. 36). Beyond a simple group-based comparison, gene content evolution was reconstructed by means of a phylogenetic birth-and-death model, which allows for predictions of functional gene content at ancestral nodes and its development along the Octadecabacter lineage. Using COGs to that end reduces the resulting data complexity, as functionally equivalent proteins are assigned to the same family, whereas the more stringent orthology detection method discussed in the previous sections might assign them to different OGs (asserting different hereditary backgrounds, even if the actual function may be similar). However, this simplification is at the cost of reduced integrity, since only 78 to 83 % of CDSs in the Octadecabacter genomes were actually assignable to a COG. Following, the general trends in COG content of polar Octadecabacters will be interpreted, including the category size changes along the Octadecabacter lineage when they are of interest (Section 4.2.1). The ecological significance of COGs gained at the polar Octadecabacter LCA and the involved pathways will be discussed in Section 4.3. Besides these results, an evaluation of the estimated parameter values and general properties of the birth-and-death model can help assess its credibility, and can reveal general trends in individual lineages. Such an evaluation will be undertaken in Section 4.2.2.

4.2.1 *General trends in COG content*

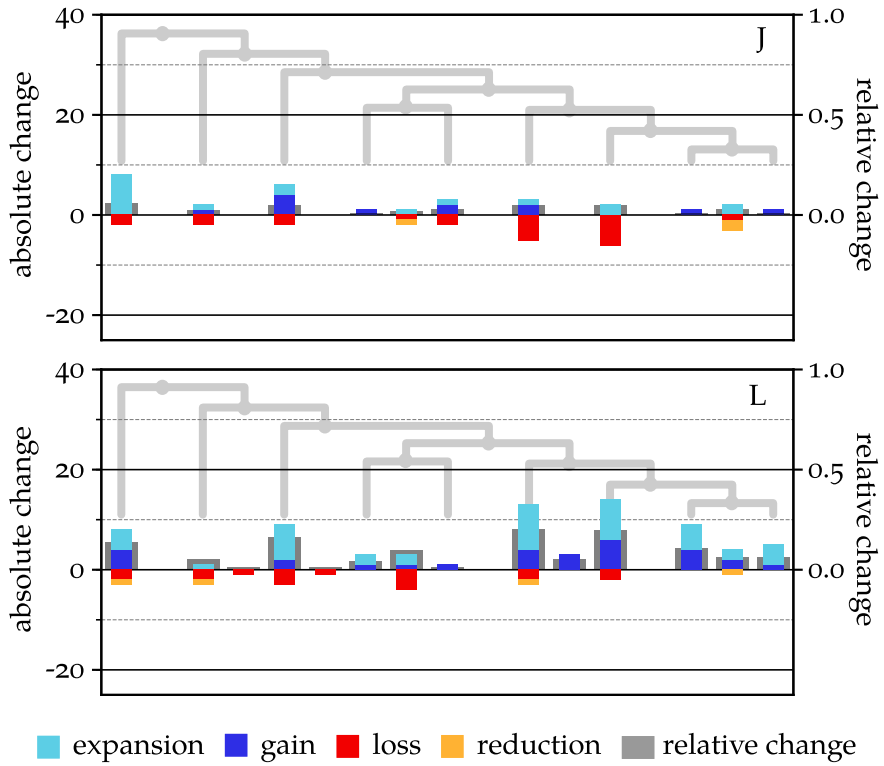
As stated earlier, polar Octadecabacters possess, on average, 800 additional CDSs in comparison to temperate isolates. These are not spread evenly among the COG categories. Genome fractions devoted to DNA metabolism and informational functions (i. e. categories J, D, F, and L) usually negatively correlate with genome size, presumably because the additional turnover generated by larger genomes can still be processed by the same number of genes in these categories [296]. This behaviour was also found in polar Octadecabacters for categories J, F, and D. Furthermore, they show little change along the Octadecabacters' ancestral lineage, judging from the probabilities assigned by the birth-and-death model (Figure 4.9).

However, COG category L (Replication, recombination, and repair) is disproportionately stronger represented in polar Octadecabacters



(a) Categories D and F

Figure 4.9: Changes in COG categories D (Cell cycle control, cell division, chromosome partitioning) and F (Nucleotide transport and metabolism) (a), and categories J (Translation, ribosomal structure, and biogenesis) and L (Replication, recombination, and repair) (b) in the Octadecabacter lineage. The number of all COGs in the indicated category, which were assigned a probability $p > 0.5$ of either gain, loss, expansion, or reduction by the phylogenetic birth-and-death model, are given above for each node in the Octadecabacter clade phylogeny (coloured bars, left scale). This number was normalised by the total number of different COGs of that category at that node (that is, all respective COGs with a summed probability $p > 0.5$ of occurring once or multiple times) to obtain the relative change, indicated by grey background bars (right scale). The overlaid cladogram indicates the order in which the comparison genomes are shown, which is based on their phylogenetic relationships. Note that the *expansion* and *reduction* bars do not represent absolute gene numbers, but only the number of COGs in which the respective change occurred.



(b) Categories J and L

Figure 4.9: (continued)

(Figure 3.7a). Its change is strongest in these species, and most of it likely occurred after branching from their common ancestor, close to the extant taxa (Figure 4.9b). This already illustrates the importance of recombinatorial effects for environmental adaptation of these strains, which will be discussed in more depth in Section 4.3.1. Notably, DNA repair mechanisms also fall into this category, which, as mentioned in Section 3.4.1, are of particular importance in polar habitats (further discussed in Section 4.3.3).

The distribution of COG categories C (Energy production and conversion) and N (Motility) in Octadecabacter genomes further contrasts previous findings. Normally they would be expected to positively correlate with genome size [296], but instead they show negative correlation[2] in polar Octadecabacters. Nevertheless, gain and expansion events dominated in both categories (Figure 4.10). All Octadecabacters possess annotated genes for flagellum synthesis, which constitute the majority of COG category N genes in their genomes. In addition, *O. arcticus* and *O. antarcticus* gained (non-paralogous, i. e. likely horizontally acquired) flagellar genes, leading to large changes in this category. Motility tests of these two strains, however, gave negative results [61]. Consequently, it remains unclear under which circumstances these two

2 term *correlation* as used in ref. [296], i. e. referring to occupied genome fraction versus genome size; see also Section 3.3.2

strains show motility, and what the function of their additional flagella synthesis genes may be.

Category C comprises functions in energy production and conversion, and includes the xanthorhodopsin specific for polar Octadecabacters, which was likely gained by a common ancestor and then passed down vertically to subsequent generations (Table 3.5). Apart from acquisition of this new function, expansion of existing gene families in this category predominate for the extant polar representatives, which contrasts more distant clade members like isolates NH9-P7 and E8 (Figure 4.10). Drastic changes concerning energy production and conversion therefore seem to have played a lesser role in polar habitat adaptation. Rather, as Sections 3.4.1 and 3.5 demonstrated, polar Octadecabacters utilise a broader range of substrates and integrate different environmental stimuli. Accordingly, COG category G (Carbohydrate Transport and Metabolism) shows a higher representation in their genomes (Table 3.4). Nonetheless, overall restructuring in this category, i. e. the amount of function gain and loss, is high in all extant clade members (Figure 4.10), likely since they stem from diverse habitats with different nutrient spectra (consider also remarks in Section 4.1). The mobilome presents a similar case, in that all extant nodes show large relative changes in COG category X (Figure 4.10), but it occupies a larger fraction in polar Octadecabacter genomes than in temperate ones (Table 3.4). This agrees with previous observations that temperate bacteria possess mobilomes which are distinct from their psychrophilic relatives, but that the latter usually contain more mobilome components, like ISs [297].

Categories R and S, comprising genes with poorly characterised or wholly unknown functions, exhibit the largest changes along the Octadecabacter lineage (Figure 4.11), although they do not constitute the largest categories overall (Figure 3.7). A high fraction of uncharacterised genes was also noted for the specific genome of Octadecabacter group II (Section 3.4.2). Consequently, a large part of functional habitat adaptations seen not only in polar, but all Octadecabacter clade genomes, is currently not available for interpretation, which is a general caveat of content-based analyses (Section 1.5).

Large genomes were found to encode a higher proportion of genes associated with transcription [296]. Polar Octadecabacters do not show this trend, although multiple transcriptional regulators are part of their group-specific genome (Section 3.4.1). Overall, the variation seen in this function (COG category K) along the Octadecabacters' lineage is of comparable degree between polar and temperate isolates (Figure 4.11). Co-factor synthesis (category H) presents a similar case, despite the polar members' increased synthesis capabilities (Sections 3.4.1 and 3.5). Both observations reflect the fact that the group-specific genome, or functional changes at the LCA, only represent a small part of the variation in genetic equipment during habitat adaptation. Rather, the ma-

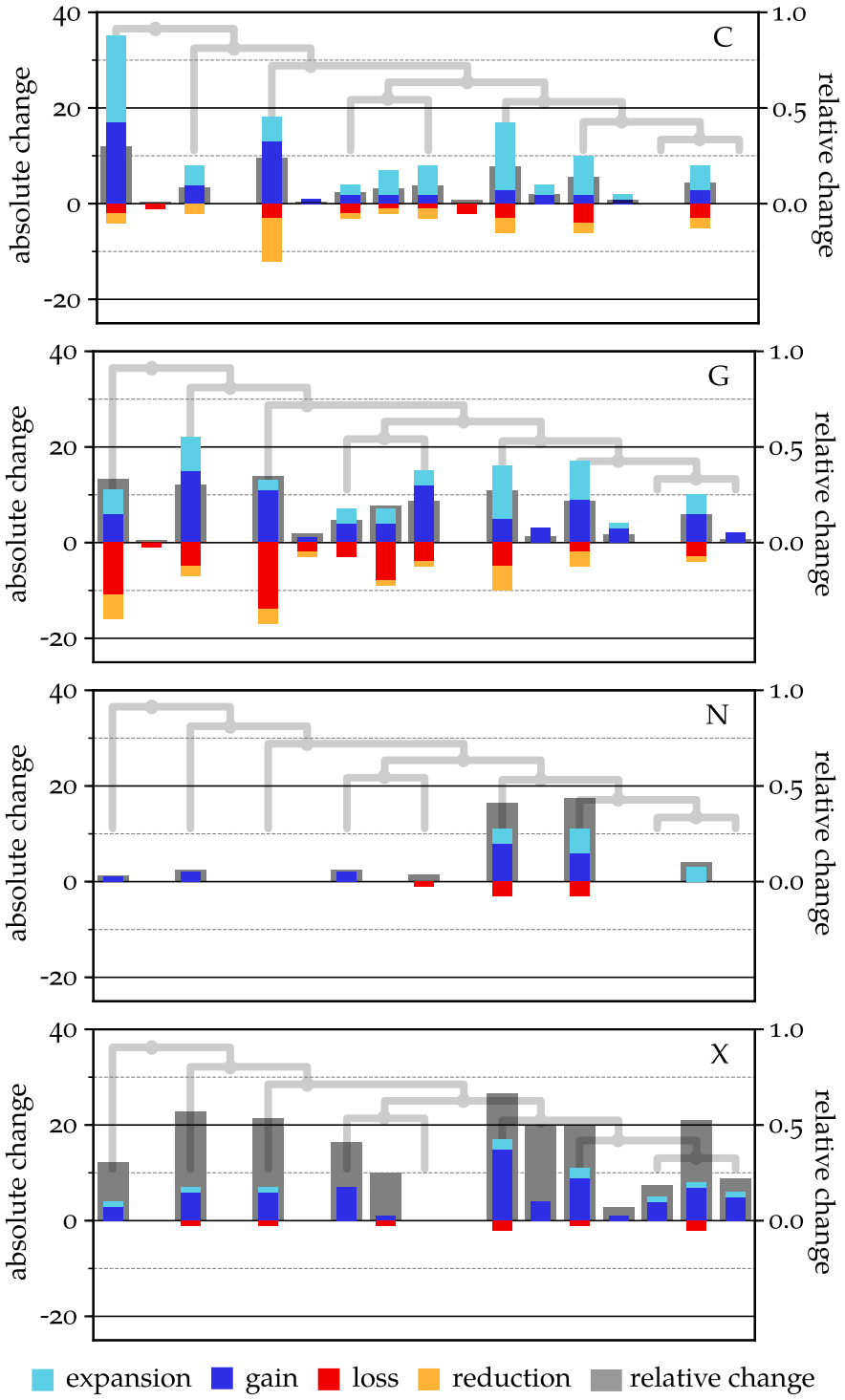


Figure 4.10: Changes in COG categories C (Energy production and conversion), G (Carbohydrate transport and metabolism), N (Motility), and X (Mobilome) in the Octadecabacter lineage. For details, see caption of Figure 4.9.

jority of changes occur on the level of the individual strain. However, group-specific genes and ancestral changes reflect more general aspects of the adaptation, independent of the ecological niche inhabited by individual polar species, and therefore allow for more generalised conclusions.

In summary, the differences in COG category distribution between polar and temperate Octadecabacters are partly derived from the former's larger genomes, particularly concerning functions in DNA metabolism and information processing. On the other hand, categories which show a stronger representation in polar genomes are more related to the ecology at this particular habitat (discussed in more detail in Section 4.3). Their increased capacity for nutrient uptake and degradation indicates a copiotrophic lifestyle, which would be plausible given the usually high microbial density and primary productivity in the sea ice habitat (Section 1.2). Earlier studies by Lauro et al. found that, compared to oligotrophs, copiotrophs possess significantly more genes involved in cellular motility, signal transduction, transcription, and defence mechanisms (COG categories N, T, K, and V, respectively), and less genes involved in secondary metabolism, and lipid transport and metabolism (categories Q and I) [298]. Judging from these observations, polar Octadecabacters do not seem to follow a more copiotrophic lifestyle than their temperate relatives, and the implications of this situation are discussed in Section 4.3.2.

4.2.2 *Properties of the birth-and-death model*

Wolf and Koonin argue that genome reduction is the predominant process in genome evolution [299]. They suggest that genome complexity does not emerge continually, but punctually, and is followed by gradual reduction as organisms adapt to their specific niche [299]. Consistently, former ancestral reconstructions in a broad phylogenetic background estimated the *Roseobacter* group's LCA to have had over 8,000 genes [93]. From this ancestor, gene counts remained high compared to the extant genomes up to relatively late bifurcations, with subsequent successive losses. According to a model computed by Luo et al., the genome size of the Octadecabacter and Loktanella clade's LCA was around 7,400 [93]. In this regard, it seems consistent that gene loss also constitutes the strongest lineage-specific component in almost all deeper branches of the model computed here (see Figure 3.19 for parameter visualisation within the Octadecabacter clade).

Luo et al. also found linear correlation between expected amino acid substitutions and gene duplication and loss rates, from which they predicted abnormally high gene duplication rates for both *O. arcticus* and *O. antarcticus* [93]. It is, however, only sensible to relate parameter values to the branch lengths in the phylogenetic ML-tree, if these correlate with the branch length parameter estimated by COUNT, for which they

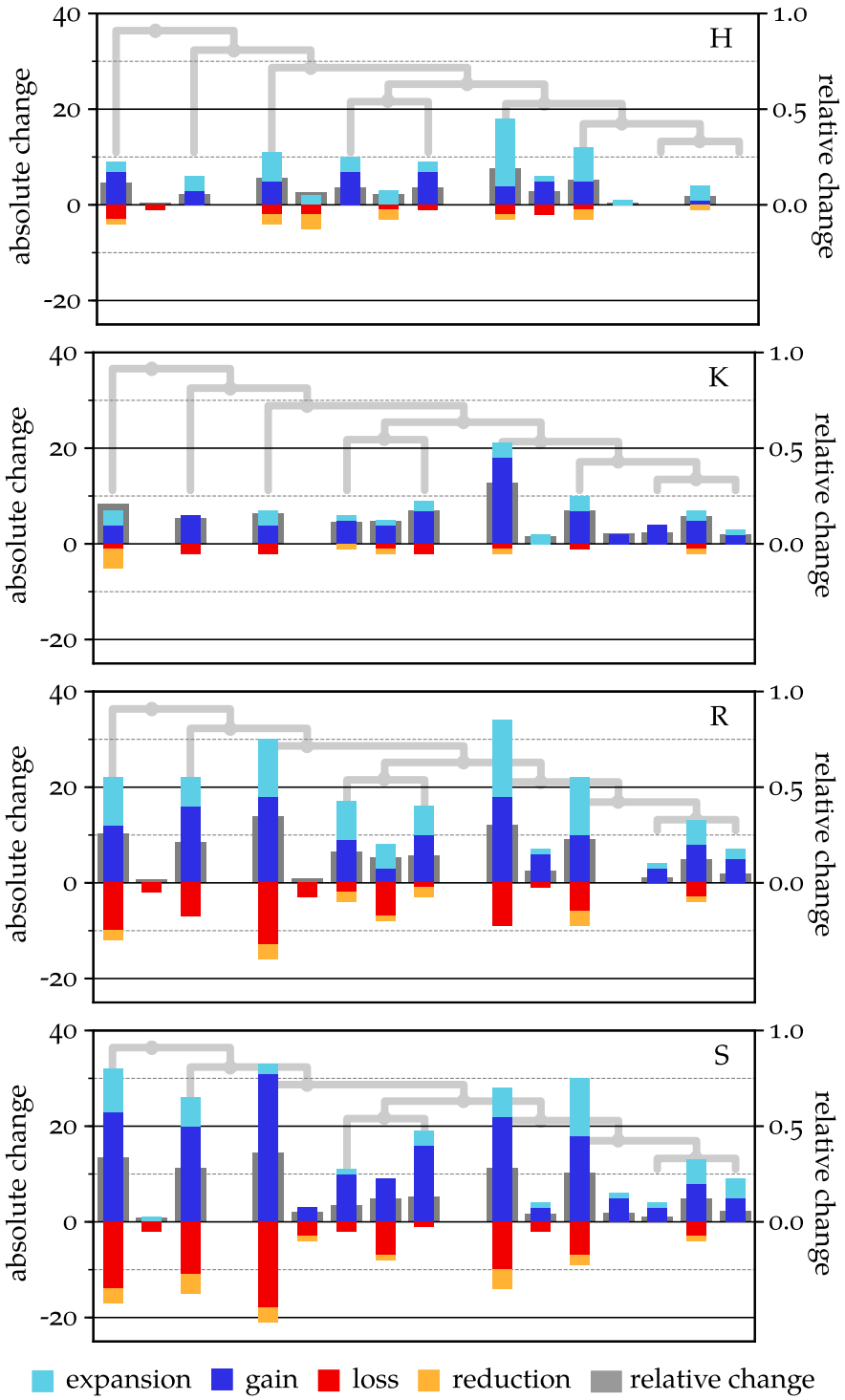


Figure 4.11: Changes in COG categories H (Coenzyme transport and metabolism), K (Transcription), R (General function prediction only), and S (Function unknown) in the Octadecabacter lineage. For details, see caption of Figure 4.9.

provide no detailed information. In the present analysis, the model tends to assign relatively shorter lengths to internal edges as compared to the ML-derived lengths, and both correlate poorly ($R^2 = 0.09$, $p = 0.82$, Figure 3.19). Edges to extant species correlate well, and show a tendency towards longer branch lengths for the polar Octadecabacters in the birth-and-death model, reflecting their more extensive content-based divergence (mentioned in Section 4.1.1). Based on these data, no investigation of clock-like parameter behaviour as in ref. [93] was undertaken. Contrary to the findings by Luo et al., the parameter values at most extant and internal nodes of the polar Octadecabacters suggest that transfer, i. e. acquisition through innovation or LGT, and not duplication, has been most important in their functional adaptation. This observation agrees with the recent finding that ≈85 % of gene family expansions are due to horizontal acquisition versus ≈15 % to duplication [289]. Only the two PAMC isolates show high duplication values, albeit at short branch lengths. The discrepancy between this and Luo et al.'s models may result from the lower depth and higher resolution of the phylogeny used here, but also from the fact that the present model was optimised on COG data, and not on orthology data derived from pairwise BLAST comparisons.

There are no striking similarities in parameter values among the three *Loktanella* species, which were isolated from Antarctic lakes [55]. *L. vestfoldensis* DSM 16212 has high rates of transfer and duplication, paralleling to some extent the polar Octadecabacters (Figure A.7b, p. 170). *L. salsilacus* and *L. fryxellensis*, which fall into a different subclade, show no strong preference of any specific parameter. This suggests that these species may have followed different strategies in adapting to their polar habitats, which will be discussed in more detail in Section 4.3.4.

Among the other genomes, parameter values along the *Ketogulonicigenium* branch strongly differ from the rest, as they possess the highest branch length and show strong tendencies towards loss of functional families (visualised in Figure A.7c, p. 171). This signifies the drastic amount of change in functional gene content associated with adaptation of their terrestrial lifestyle.

4.3 Environmental adaptations in polar Octadecabacters

Sections 3.4 and 3.5 described characteristics of polar Octadecabacters and provided basic interpretations on their functions and roles in these organisms. This section seeks to integrate these data into a comprehensive view of how polar Octadecabacters adapted to their environment. This includes analysing and rating their basic mechanisms of adaptation (Section 4.3.1), and apparent methods to deal with higher levels of stress (Section 4.3.3). Several metabolism-related content alterations

were found, and these are interpreted in Section 4.3.2. Finally, Section 4.3.4 compares characteristics of polar Octadecabacters to other polar isolates within the *Roseobacter* group and discusses possible reasons behind the observed similarities and differences.

4.3.1 *Genomic adaptation mechanisms*

Colonisation of a new habitat presents microbes with the challenge of adapting to new selective stimuli, potentially encompassing extreme environmental situations. Adaptation takes place via three mechanisms:

i changes in gene regulation, sometimes induced by genomic rearrangements, which alter regulatory networks,

ii protein family expansion resulting from LGT, or from duplication of existing genes, and

iii point mutations, which confer a selective advantage and therefore become fixed in the population.

The data presented in Sections 3.4 and 3.5 show that functions associated with these mechanisms are more abundant in polar Octadecabacters than in their temperate relatives, which is a trend often observed in psychrophilic bacteria [297].

In addition to GTAs, for which at least partial gene clusters are present in all currently sequenced Octadecabacters, the high number of transposable elements (TEs) and functions associated with recombination (Section 4.2.1) facilitates not only genomic rearrangement, but also uptake of laterally transferred genetic material [300]. In Section 3.4.1 some genetic loci were described, which likely constitute parts of regulatory networks that are specific to polar Octadecabacters. Many of the corresponding functions coded therein are not per se exclusive to this group, such as the carbohydrate ABC transporters, or are represented by additional orthologues elsewhere in the genomes, e. g. the TCA cycle-associated dehydrogenases. Since these are not paralogous, but constitute individual OGs with equivalent functions, they were likely acquired horizontally in the ancestral lineage of polar Octadecabacters. This would also explain their conserved order in the four genomes. The conservation of these particular loci, despite the high rate of rearrangement in these genomes (demonstrated for *O. antarcticus* and *O. arcticus* in fig. 3 of ref. [77]), furthermore indicates that they are operons of functionally correlating genes. Consequently, LGT seems to have played a role not only in acquiring completely new functions such as the xanthorhodopsins, or DNA polymerase V, but also in expanding functionalities already present in the genomes.

Regarding protein family expansion, LGT is markedly more frequent in situ than duplication [289]. Nonetheless, duplication allows for relatively fast adaptation within the time frame of a few generations [301].

Polar Octadecabacters generally seem to have undergone more stable gene duplication events than their closest relatives, as indicated by their higher average number of paralogues (see Table 3.6). This likely results from their higher number of mobile genetic elements, i. e. transposases, ISs, etc. At this point, it is worth noting that the activity of such elements is often increased in situations of stress, demonstrated for, e. g. oxidative stress [302], starvation [303], or UV-induced damage [304]. Multiple forms of stress are frequently encountered at polar habitats (and sea ice in particular, as explained in Section 1.2), and a large mobilome is advantageous to polar Octadecabacters, as it increases their evolvability under these conditions.

Various group-specific genes discussed in Section 3.4.1 are found in multiple copies within individual polar Octadecabacter genomes, e. g. DNA polymerase V, NfuA, or pyridoxamine 5'-phosphate oxidase, underlining their apparent importance for habitat adaptation. Nonetheless, there is no polar group-specific gene, which is duplicated in all four members of this group. The high fragmentation of the two PAMC isolate genomes may partially obscure their actual gene content, and therefore lead to an underestimation of their TEs and paralogue numbers (see also Appendix A.7, p. 203).

Generally, 45 to 92 % of paralogues are exclusive to each strain, with the exception of the two highly similar PAMC isolates, and predominantly affect the cloud-, and shell-genomes (Table 3.6). Interestingly, isolate NH9-P7 possesses a number of paralogues similar to that of polar Octadecabacters, despite its lower number of TEs (Figure 3.7). The majority of its duplications also occurred in the clade's core-genome. Consequently, the number of mobile genetic elements alone is not necessarily a direct indicator for the number of duplications accumulated in a genome. A larger number of these elements will increase the capacity for short-term changes, which are stimulated by conditions of stress, as discussed above.

Notably, the results of transposition and homologous recombination events of mobile genetic elements are indistinguishable [300]. It is therefore unclear, which of these processes dominated in polar Octadecabacters.

TEs not only mediate gene duplications, but also horizontal transfer and rewiring of genetic networks [305]. In that function, they constitute one specific strategy an organism can follow in order to adapt. This tactic in some cases trades speed for efficiency, as mutations generated by TE activity may hinder the fixation of other, more beneficial mutations, e. g. point mutations in specific promoters [306].

Compared to TE-derived mutations, point mutations are a slower adaptive process, usually taking thousands of generations [301]. Their rates are increased in polar Octadecabacters, likely due to the acquisition of DNA polymerase V. As mentioned in Section 3.4.1, this enzyme

causes a higher mutation rate while bypassing and repairing DNA damage, thereby increasing evolvability [230].

Sea ice is thought to be a "hot spot" of microbial evolution due to its high abundance of phages and lower selective pressure due to grazing [307]. No phage-associated genes were found specific for polar Octadecabacters, and it is unclear how much of their mobilome, e. g. specific transposases. or specific gene content results from acquisition of phage DNA. Transfer and uptake of DNA is however also facilitated by the other mobilome components. Importantly, the frequent encounters of stressful situations in sea ice, e. g. rapid shifts in salinity, or phases of heavy metal exposure, require and will therefore select for highly mutable genomes. Many components of the mobilome, e. g. specific transposases or DNA polymerase V are specifically activated during stress response. This system of reoccurring stress signals and concomitant mutagenesis constitutes another inherent reason why this habitat acts as evolutionary "hot spot".

4.3.2 *Metabolic properties of polar Octadecabacters*

Polar Octadecabacters have acquired the ability to utilise a broader range of substrates for metabolism than other members of this clade, some of which have been mentioned in the previous sections (e. g. histidine, fuculose/arabinose). More complex metabolic networks usually require more complex regulation, and both features are associated with larger genomes [296, 308], as observed for polar Octadecabacters (Table 3.1). Metabolic complexity also correlates with the high rates of genomic rearrangement present in some polar Octadecabacters [77], since in larger genomes with more regulators, natural selection for operons is relaxed [309].

Compared to the marine environments from which other Octadecabacters were isolated, this difference in metabolic capabilities likely results from sea ice being a complex, highly productive habitat [42] where microbial cell densities may reach over 10^7 cells/mL [310]. The sea ice residents *O. arcticus* and *O. antarcticus* therefore likely live in close association with primary-producing *Diatoms*, as it is common for the most abundant sea ice bacteria [311]. Such relationships usually influence the capacity for vitamin and co-factor synthesis, and the data presented throughout Sections 3.4.1 and 3.5 indicate that polar Octadecabacters act as providers of thiamine (Section 3.5.1) and cobalamin, which are frequently exchanged between prokaryotes and their algal hosts [256, 312]. Nonetheless, when cultivated in vitro, *O. antarcticus* and *O. arcticus* cannot grow without vitamin supplements, indicating auxotrophy for at least one co-factor [76] (for example pantothenate, as noted in Section 3.5.6).

The fact that most degrading enzymes require co-factors for their reactions may also link their increased production capabilities to the

broader range of degradable substrates in polar Octadecabacters. To name one example, the glucan phosphorylase activity likely gained at the polar Octadecabacter LCA (Section 3.5.6) requires pyridoxal-phosphate, which all Octadecabacters can synthesise. The polar group gained an additional, group-specific pyridoxamine 5'-phosphate oxidase for its synthesis (Section 3.4.1).

In their relation with primary producers, the associated prokaryotes gain access to degradable photosynthates in exchange for their secreted co-factors, and this relationship is the main carbon source for heterotrophic bacteria, as detailed in Section 1.1. Indicative of this relationship are the glucan phosphorylase and glucanotransferase gained by the polar Octadecabacter LCA (Section 3.5.6), as well as their increased carbohydrate transport capabilities mentioned in Section 3.4.1. Due to the high levels of primary production [42], and since EPS produced by *Diatoms* also serve to concentrate DOC [37], sea ice generally favours a copiotrophic lifestyle. It is generally accepted that copiotrophs possess more genomic copies of the rRNA operon [298, 313], as well as larger genomes and increased transport and degradation capabilities [298]. Polar Octadecabacters exhibit these features in comparison to the other clade members, although the higher rRNA operon copy number is masked in the PAMC isolates by their high fragmentation (but see Appendix A.7, p. 203). Nonetheless, they do not show the typical COG profiles associated with copiotrophy (Section 4.2.1). They furthermore exhibit slow growth rates even in full medium (for data on *O. arcticus* and *O. antarcticus* see ref. [61]), which is more typical of oligotrophs. One explanation might be that they require parts of their increased metabolic capabilities and available energy for the repair of cellular damage, at the cost of growth speed and efficiency. A higher number of rRNA and tRNA genes was also suggested to compensate for reduced translation speeds at low temperatures [39]. Notably, numerical dominance of α-Proteobacteria, and Octadecabacters in particular, seems to be relatively uncommon in sea ice, and other known copiotrophs from the γ-Proteobacteria and Bacteroidetes usually dominate the prokaryotic communities [53, 311].

It could seem paradoxical that the sediment-residing PAMC isolates, which were procured from a depth of 156 m [80], show the same metabolic and substrate patterns as the sea ice residents *O. arcticus* and *O. antarcticus*. Studies in the Arctic showed that during melting of sea ice, algal biomass sinks to the ocean floor and can reach depths over 4,000 m [314]. Such a process could avail the two isolates the same nutrient spectrum as their surface-dwelling relatives.

In contrast to sea water, nitrogen is abundantly present in sea ice, and does not usually constitute a limiting factor for prokaryotic growth [53, 315]. This is reflected in the group-specific Octadecabacter genomes, since only group III possesses multiple extracellular proteases, which are secreted to make nitrogen accessible through the degradation of

extracellular proteins. Thus, in concordance with the presence of a specific starvation-inducible factor, this group seems to frequently encounter phases of (nitrogen) starvation. On the other hand, the number of genome-encoded metalloproteases seems to have undergone a reduction in the polar Octadecabacters' ancestral lineage (Table 3.5).

Rather than nitrogen, phosphorus seems to be the most limiting element in sea ice [53]. However, it is unclear from the group-specific genome, in how far this influences the core metabolism of polar Octadecabacters. Curiously, a nucleotide phosphodiesterase likely experienced family reduction along their ancestral lineage, which would make inorganic phosphate accessible from extracellular nucleotides (Section 3.5.6). Conversely, a potentially higher amount of rRNA, implied by the increased rRNA operon copy numbers in this group, would constitute a larger sink of cellular phosphorous [316]. The presence of TauD may cause a slight alteration in their inorganic phosphate pool management (see below).

Substrate utilisation is generally controlled through the integration of environmental stimuli, and most polar group-specific loci related to degradation are preceded by genes encoding transcriptional regulators (Section 3.4.1). In addition, polar Octadecabacters possess more genes coding for enzymatic functions which steer cellular metabolism. For example, as mentioned in Section 3.5.6, GPDH regulates substrate flow between fatty acid synthesis and glycolysis, and this functionality expanded in this group. One function in this regard, which is absent in all other Octadecabacters, is the taurine dioxygenase TauD. Taurine is abundant in the environment, and serves as a nitrogen, carbon, and sulphur source [317]. All Octadecabacters are genetically able to degrade it via the *tpa*/*xsc* route, which first assimilates nitrogen (*tpa*), and then sulphur (*xsc*) from taurine, ultimately producing acetylphosphate [317]. TauD, on the other hand, directly obtains sulphite from taurine in an oxygenolytic reaction [238]. Importantly, it also acts on a number of different substrates other than taurine [238]. In *E. coli*, TauD expression is induced under sulphur-limitation [238], and *Rhodococcus opacus* explicitly does not use the *tpa*/*xsc* pathway for sulphur assimilation from taurine, but likely uses TauD [318]. Polar Octadecabacters may follow a similar tactic, since sulphur-assimilation via TauD has the advantages of i) consuming oxygen, thereby potentially reducing the level of oxidative stress, as mentioned in Section 3.4.1, and ii) conserving phosphate in the cell, as no acetylphosphate is produced. Thus, TauD is an example of a newly acquired function in polar Octadecabacters, which acts at the intersection of multiple cellular element cycles (N- and S-cycles), and may therefore serve a specific regulatory function in this group. The easy acquisition of sulphite from this abundant substance may explain why this group has no need for the methionine salvage pathway (Section 3.5.3). Most Octadecabacters, including the polar group, are furthermore genomically able to degrade the sul-

phur compound 2,3-dihydroxypropane-1-sulfonate (DHPS), which is produced and secreted by most algae and and a known "currency" in their symbiotic relationships with roseobacters [65].

The more immediate uptake and therefore regulatory capability of the S-cycle by TauD may also be related to the increased stress levels at the polar regions. Section 3.4.1 mentioned that the oxidative state of sulphur atoms in biomolecules is particularly vulnerable to modifications by free radicals, and the products of various group-specific genes, such as NfuA, or the thiol-disulfide oxidoreductase, relate to alleviating such damage. The low molecular weight sulphur compound glutathione is highly abundant in α-Proteobacteria and is one of the most important cellular antioxidants [319]. Even under growth conditions without the presence of external stress agents, glutathione turnover in the cell is higher than that of most co-factors [320]. With turnover potentially increasing in their more oxidative environment, polar Octadecabacters may face higher loss of sulphur species due to irreversible damage, or need to rapidly synthesise quantities of glutathione (or other antioxidative sulphur compounds) during phases of acute oxidative stress. TauD would allow for a more rapid mobilisation of sulphur from taurine, or similar molecules, than the *tpa*/*xsc* pathway, and would therefore be beneficial under such circumstances.

A further notable aspect is their phototrophic capability due to presence of a xanthorhodopsin system (Section 3.4.1). Phototrophy is frequently encountered in marine habitats, since it compensates for the general lack of nutrients in most oceanic water masses, or improves trophic efficiency [205, 251]. Of the two known alternatives, photochemical reaction centres are the more complex one, requiring 30 to 40 gene products for functional assembly [321], and forming huge membrane-situated complexes [322]. Since membrane biochemistry and protein-protein interactions are significantly altered in cold environments [28, 323], it is plausible that the polar Octadecabacters favour the more simple (xantho)rhodopsin system, which consists of one protein and its chromophor, and requires five genes for its biosynthesis. Furthermore, the activity of photochemical reaction centres generates significant amounts of oxidative stress [252]. The higher oxygen solubility at low temperatures, and the enrichment of heavy metals at both poles [46–48] both generate high background levels of oxidative stress, which might make the additional cost of reaction centres too high. In contrast, as discussed in Section 3.4.1, rhodopsins exhibit a protective role during oxidative stress by maintaining the membrane potential during disturbances in the quinol-quinone-cycle, in addition to their potential role during starvation. Nonetheless, photochemical reaction centres associated to *Roseobacter* group bacteria also occur in sea ice, but at lower frequency than in the underlying sea water [90]. Exclusively favouring rhodopsin-based phototrophy therefore constitutes a characteristic of

polar Octadecabacters (discussed in more detail in Section 4.3.4), although its function in these species remains unknown.

During phases of high primary production, degradable substrates are available in copious amounts. Their production shows high spatiotemporal variability. For example, glycerol synthesis is stimulated during phases of osmotic stress. At the same time, the sea ice can melt and release its microorganisms into the water column, thereby potentially disrupting the syntrophic associations and greatly reducing substrate availability. The group-specific genome of polar Octadecabacters contains several indicators that these strains repeatedly have to deal with such phases. For example, rhodopsin activity has been implicated as a mechanism to resist starvation [205]. In addition, cobalamin-independent methionine synthase constitutes a redundant enzymatic function without the requirement of a (costly) co-factor. Finally, as mentioned in Section 3.5.1, this group possesses more functions related to thiamine degradation and recycling, which could be related to a more efficient utilisation of available cellular compounds.

Interestingly, all Octadecabacters are also capable of anaplerotic carbon fixation through the malic enzyme route (explained in ref. [324]), which may reduce their reliance on exogenous substrates.

In conclusion, the specific genome and enriched functions of polar Octadecabacters reflect metabolic adaptations, which are consistent with our current knowledge of sea ice ecology (see above). Like other studied members from the *Roseobacter* group [54, 257], these strains seem to be closely associated with primary producers in a mutualistic relationship.

The present analysis cannot provide statements as to how polar Octadecabacters interact with the other abundant and usually copiotrophic prokaryotic sea ice residents. While only substrate specialisation and concomitant mutualism theoretically lead to a collective optimum of biomass [325], it remains unclear at this point whether cooperative or competitive interactions dominate the relationship among these members of the sea ice microbial community (SIMCO).

It is noteworthy that polar Octadecabacters are the only group, the specific genome of which renders a relatively complete and interpretable picture of metabolic adaptations, as the above discussion of additionally utilised substrates and environmental integration showed. This may be due to their high homogeneity (i. e. pairwise similarity) in comparison to the other groups (Figure 3.1). Furthermore, most of their specific genes are functionally characterised. Conversely, the specific genome of *O. temperatus* and *O. ascidiaceicola* contains mostly uncharacterised genes. This group would thus provide a good basis for further experimental study, as their characterisation could illuminate the lifestyle aspects which both strains share, and which seem to be poorly covered by our current knowledge.

4.3.3 *Genomic indicators of psychrotrophic adaptations*

Cold habitats constitute a large part of earth's biosphere, and cold adaptation in bacteria has frequently been studied on the genomic level (Section 1.2). Fluidity of the cellular membrane is usually maintained at low temperatures by increasing the amount of unsaturated fatty acids, and fatty acid desaturases were identified as specifically enriched in some cold-adapted organisms [15]. This is not the case in polar Octadecabacters. They possess multiple desaturase genes, but none of these is specific to this group. The average number of desaturase genes is furthermore not higher for polar Octadecabacters than for the temperate strains.

Existing data, which were compiled by Billerbeck et al., also show that the lipid profiles of the compared clade members are relatively similar, and that the content of polyunsaturated fatty acids in temperate isolates can actually exceed that of polar ones at low temperatures (table 2 in ref. [79]). Polar strains seem to possess a slightly higher percentage of shorter fatty acids, as it is common for cold-adapted organisms (Section 1.2). Some adaptation may also derive from differential regulation of fatty acid synthesis, since polar Octadecabacters possess additional GPDH functionality (Section 3.5.6), which can steer substrates between fatty acid synthesis and glycolysis. Beyond these relatively minor alterations, polar strains seem to counter the altered membrane physiology through other means, such as an increase in degradation capabilities (indicated by the group-specific AAA ATPase).

In addition, polar Octadecabacters lack group-specific antifreeze proteins (AFPs) or cold-shock proteins (CSPs), which are proteins that usually accompany cold-adaptation in bacteria (Section 1.2). Conversely, the temperate group III possesses one specific CSP. As mentioned in Section 3.4.3, CSPs are particularly important during the cold-shock response in mesophiles, where they halt translation and regulate transcription, in order for the organism to adapt to the new temperature [30]. Stenopsychrophiles like the polar Octadecabacters possess a constitutively cold-adapted transcriptional and translational machinery, and their CSPs rather serve for cold-acclimation [27]. Due to this lesser role, it is plausible that this group does not require additional CSP functionality for efficient growth. Cold-adaptation of the translational machinery in polar Octadecabacters is evidenced by the group-specific ribosomal protein RpsU (Section 3.4.1) and, as mentioned in the previous section, probably by the higher number of rRNA gene copies. For the replicative machinery, gain of a DNA replication protein (COG 1484, Table 3.5) seems to constitute one adaptive trait (further discussed in Section 4.3.4).

With regard to ice crystal formation, the lack of AFPs indicates that polar Octadecabacters rely on other methods to avoid water freezing

and its accompanying effects. For example, they possess increased capacity for the production of proline (Section 3.5.5), and for glycerol import (Section 3.4.1), which are both effective compatible solutes. Likewise, the presence of genes encoding exopolysaccharide synthesis and export proteins suggests that all members of this group are able to produce extracellular polysaccharides (EPS) themselves, and therefore likely do not need to rely on EPS produced by *Diatoms* to repress the freezing point of water.

Enzymatic function is usually maintained at low temperatures via a decrease of protein stability, for which exchange of only a few amino acid residues (<1 %) can suffice [326]. Since secondary structures shape the tertiary structure of a protein, and are therefore essential for its function, it is thought that such features do not differ much between temperate proteins and their cold-adapted homologues [29]. This seems to hold true for the core-proteome of Octadecabacters. Concerning relative secondary structure content, polar strains only form a distinguished group when both structural features, α-helix and β-sheet, are considered in combination (Figure 3.21). This clustering may simply result from the fact that they are more similar to each other on the sequence level than to the rest (Figures 3.1 and 4.7). Overall, for them there is clearly no significant trend towards a higher, or lower structural feature content (Figure 3.22).

Observed differences in amino acid frequencies between psychrophilic and mesophilic microbes can vary largely, depending on the organism's class, GC-content, or the phylogenetic distance to and nature of the mesophilic comparison organisms [27, 35, 86, 327]. Particularly GC-content is a strong determinant of amino acid profiles [328], which is why *P. jejudonensis* and isolate NH9-P7 were excluded from the comparison of amino acid frequencies within the Octadecabacter clade (Table 3.1). When compared to the remaining three temperate strains, polar Octadecabacters show only slight deviations in amino acid frequencies (Figure 3.23). Typical observations, such as reduced arginine content [35], or increase in glycine, and decrease in proline content [28] in psychrophiles, are not, or only weakly (glycine) evident in this group when the whole core-proteome is considered. However, they show higher frequencies of some hydrophobic amino acids (four out of six, Figure 3.23), a trend which is also present in other psychrophiles [35]. Moreover, these differences reflect to some extent correlations between amino acid content and optimum growth temperature noted for some model proteins. Specifically, the amino acids which occur more frequently in polar Octadecabacters tend to be less hydrophobic than the others within the same group (compare data compiled in table 2 of ref. [329]), and hydrophobicity was found to positively correlate with growth temperature [329]. The average molecular weight presents a similar case, with cold-adapted enzymes preferentially incorporating lighter amino acids [329]. In polar Octadecabacters, this trend is less

pronounced than the decrease in average hydrophobicity. Yet, it is still notable, particularly in the group of hydrophobic amino acids, where the frequency of light amino acids, glycine and alanine, is higher in polar strains than in the temperate ones. Thus, the differences in amino acid frequencies between polar and temperate Octadecabacters are broadly consistent with general trends in cold-adapted enzymes, but agree less with data from direct comparisons of other psychrophilic and temperate organisms.

The frequency variation of an individual amino acid, which accompanies a protein's stability reduction, can be interpreted on the basis of its flexibility, and accordingly its rigidity, as captured by atomic displacement parameters, or *B-factors* [330]. Overall amino acid frequencies in polar versus temperate Octadecabacters do not display a tendency towards higher flexibility, judging from their B-factors as determined by Smith et al. [330] and compiled in ref [29] (see Figure 3.23). Conversely, such a tendency was shown for psychrophilic enzymes from other organisms [331]. However, α-helices of polar Octadecabacters show a consistent trend of avoiding rigid amino acids (I, M, and A), and increasing their content of more flexible ones (mostly D, E, see Figure 3.24). Organisms can further modulate helix stability particularly through the modification of charged amino acid numbers at the helix caps [29]. A breakdown of amino acid frequency by individual sub-domains and structural sub-features is beyond the scope of this analysis, but the fact that both negatively charged amino acids (and one out of three positively charged ones) deviate in frequency in polar α-helices (Figure 3.24) could imply the presence of this type of modification. Thus, indications are strong that polar Octadecabacters selectively reduce stability of their α-helices.

Rather than modifying their existing enzymes, bacteria often acquire isozymes, which function better at lower temperatures, as part of their cold-adaptation strategy [332, 333]. Some of the group-specific genes described in Section 3.4.1 are homologous versions of functions contained in the Octadecabacter core-genome, such as *myo*-inositol dehydrogenase, or the cluster of specific dehydrogenases. These could well constitute cold-adapted enzymes, which polar Octadecabacters acquired through lateral transfer as alternatives to those coded in their core-genome. However, due to lack of experimental data, it cannot be ruled out that they simply integrate different environmental stimuli as part of these strains' expanded metabolic networks, or serve both of these described purposes. Nevertheless, the higher capacity of polar Octadecabacters for lateral acquisition of genes, or, more generally, their high genome flexibility discussed in Section 4.3.1, thereby also constitutes a prerequisite for efficient cold-adaptation. Similar observations were also made for *Photobacterium profundum*, where inactivation of transposases lead to a cold-sensitive phenotype [334].

Beside the traits of cold adaptation discussed so far, resistance to (oxidative) stress is prominently featured in the polar group-specific genome and in this group's expanded functionality (Sections 3.4.1 and 3.5.5).

Some of the involved genes are part of the SOS-response. As mentioned in Section 3.5.5, this response is induced by DNA damage and leads to expression of multiple genes, mediated by the cleavage of the transcriptional repressor LexA [276]. The precise number and nature of genes under LexA-regulation varies between taxa, but its core-regulons in different bacterial groups usually include the *lexA* gene itself, transposases, helicases (such as RuvABC), and DNA polymerases capable of translesion DNA synthesis (TLS) [276]. Polar Octadecabacters possess more copies of LexA, transposases, and a specific copy of DNA polymerase V, and therefore show increased functionality to cope with DNA damage, as conveyed by the SOS-response. The higher number of *lexA* genes could mean a higher base expression, or a more rapid induction of this repressor, and may be particularly important for the survival of this group when one considers the function of DNA polymerase V: this enzyme, in addition to mediating TLS, acts as DNA damage checkpoint and can halt DNA synthesis, depending on the state of the UmuDC complex [335]. The latter function leads to a cold-sensitive phenotype in *E. coli* when UmuDC levels become too high [336]. Thus, both to prevent the rapid accumulation of random mutations, as well as to enable cell cycle progression, UmuDC expression has to be tightly controlled in polar Octadecabacters. Moreover, this control has to take place against high background levels of DNA-damaging agents (Section 1.2), which may necessitate more of the transcriptional repressor LexA.

Increased levels of (oxidative) stress also lead to more unwanted modifications of metabolites, which the cell has to deal with. As noted in Section 1.5, we currently only have rudimentary understanding of this particular aspect of metabolism. Nonetheless, some of the polar Octadecabacters' characteristics are likely associated with this effect. For example, the high number of paralogues in their genomes can be useful to deal with metabolites that are chemical variants of normal compounds [95], e. g. irregular variants derived from (unwanted) radical reactions. On the other hand, due to the high diversity and productivity of their environment, this could also simply result from a high substrate variation in their habitat. It would be interesting to experimentally test and compare, which of these purposes outweighs the other, both in the polar group, as well as isolate NH9-P7, which shows similarly high paralogue numbers (Table 3.6).

The modification of sulphur species in oxidative reactions, and the concomitant increase in polar Octadecabacters' respective damage protection and repair capabilities was detailed at multiple places in Chapter 3. At the same time, it is of equal importance to guarantee the integrity of other cellular compounds such as nucleotides, or co-

factors. Nudix hydrolases are often involved in quality control of the former [95], and this functional family expanded at the polar Octadecabacters' LCA (COG 1051, Section 3.5.5). Involvement of this family in dealing with oxidatively modified compounds could also explain their role in heavy metal resistance detailed in Section 3.5.5. Besides this example, multiple uncharacterised oxidoreductases were likely gained in the ancestral lineage of polar Octadecabacters (Table 3.5), and such enzymes are often part of mechanisms which repair irregular chemical modifications of molecules [95].

Recent experimental data indicate that the rate by which most co-factors are synthesised is adjusted to account for the dilution resulting from cell division [320]. Conversely, this dilution accounts for most of the co-factors' turnover, as cells minimise their loss due to unwanted and damaging (e. g. oxidative) reactions via specialised repair mechanisms [320, 337]. The fact that the cultivated polar Octadecabacter species have long doubling times implies a decreased turnover of most co-factors in these organisms and a concomitantly higher importance of their repair mechanisms. Importantly, organisms counter co-factor damage by degrading and re-synthesising them from their scavenged components. Apart from the metabolic advantage of potentially conserving nutrients, the additional capacity for thiamine synthesis and degradation present in polar Octadecabacters (Section 3.5.1) may therefore also play a role in their management of damage to this compound through unwanted modifications. This may also apply to the alterations in their biosynthesis machinery of other co-factors, detailed e. g. in Sections 3.4.1 and 3.5.2. Yet, without experimental verification, it remains unclear to what extent these relate to guaranteeing the proper chemical nature of the respective co-factor, or to their generally changed biosynthesis patterns due to different syntrophic relationships discussed in the previous section.

It was noted earlier that besides passively reacting to oxidative stress through the mechanisms described in the preceding paragraphs, polar Octadecabacters also possess more enzymatic functions which consume oxygen, and could thus actively eliminate an important source of this type of stress. Interestingly, epiphytic bacteria on a sea ice *Diatom* were shown to reduce its oxidative stress levels resulting from photosynthesis [54]. Among these bacteria were *Sulfitobacter* species, which belong to the *Roseobacter* group. Since polar Octadecabacters also seem to be closely associated with sea ice primary producers, they may plausibly be involved in the same kind of relationship, and this would be interesting to test experimentally.

One particular feature of sea ice is its high potential for osmotic stress (Section 1.2). The specific genetic equipment of polar Octadecabacters contains several ion transporters, mostly Na^+/H^+ antiporters, which are implicated in osmoregulation (Sections 3.4.1 and 3.5.5). Importantly, such transporters are also frequently associated with cold-

adaptation [39, 338]. As mentioned in Section 1.2, cells may increase their cytoplasmic ion concentrations to suppress the freezing point of water, and these transporters may therefore play a role in maintaining the balance between ion influx and export. It was also noted that the polar Octadecabacters' transport capability for the common polyamines spermidine and putrescine is increased as well (Section 3.4.1). Apart from acting as compatible solutes, i. e. osmoprotectants, they also constitute nitrogen reservoirs and protect DNA against oxidative damage. Despite this range of desirable functions, and the likely absence of Nitrogen-limitation (see previous section), polar Octadecabacters cannot synthesise these compounds on their own. This may be due to the fact that spermidine is toxic at low temperatures, likely by inhibiting ribosomal activity [339]. Rather, as mentioned earlier, they seem to prefer proline and glycerol as compatible solutes.

4.3.4 *Comparison to other polar members of the Roseobacter group*

Roseobacter group bacteria seem to be well represented in most polar aquatic environments, such as polar waters [89], sea ice [51, 90], Antarctic lakes [55], or marine sediment [80]. Comparison of key features between polar and temperate isolates of three clades other than the Octadecabacters indicated that the trends observed in the latter are not common in polar-adapted roseobacters (Figure 3.20). More precisely, only polar *Loktanella* species showed a potentially increased capacity for genetic mobility (i. e. more genes in COG categories L and X), and above-median numbers of singleton genes and paralogues, as found to be characteristic for polar Octadecabacters. The differences in these key characteristics between polar and temperate clade members are smaller for the Loktanella clade compared to the Octadecabacters. The single polar isolates associated with the *Leisingera* and *Sulfitobacter* genera, respectively, did not display these trends at all (Figure 3.20).

Consequently, it seems that large-scale increase in genomic flexibility as an adaptive strategy to polar habitats is a characteristic of the Octadecabacter clade, and polar members of other clades follow other routes of adaptation. This increase is in agreement with general observations regarding sea ice habitats [307], but can also be observed in the sediment-residing PAMC isolates, indicating that it is not specific for one particular type of polar habitat, but rather for this group of organisms as a whole. The polar *Loktanella* strains, on the other hand, follow a markedly different lifestyle (i. e. bio-mat association), which may require other evolutionary strategies for efficient adaptation.

The latter were also included in the phylogenetic birth-and-death model, which enabled a comparison of gain and loss of specific genes along both their and the polar Octadecabacters' ancestral lineages (Section 3.5). Two functions were specifically acquired along all polar lineages: COG 1484 (DNA replication protein DnaC), and COG 2608 (Copper

chaperone CopZ). CopZ is a copper-efflux chaperone, and therefore plays an important role for the cell in reducing oxidative stress [264]. DnaC acts as a regulator of the DNA helicase DnaB, and is essential during replication initiation [340]. While the former is clearly involved in a function that is of higher importance in cold than in temperate habitats, the meaning behind the gain of DnaC in a polar context remains enigmatic. More precisely, due to its essential function, all comparison strains possess an annotated version of it. However, the respective gene shows conflicting orthology relationships: one orthologous group (OG) shared by all is annotated as DnaC in some strains, and as DnaB in others, whereas it is assigned to COG 0305 (Replicative DNA helicase). COG 1484, on the other hand, occurs as multiple, mostly singleton, OGs in the polar isolate genomes, and was assigned as either DNA replication protein, or as an insertion sequence (IS)-associated protein. Consequently, the current unknown functions and relationships of these genes bar further insight at this point, but they could be interesting candidates to study their function and potential role in psychrotrophic adaptation in vivo.

As noted in Section 3.4.1, rhodopsin phototrophy seems to be frequent in polar *Roseobacter* group isolates, and it constitutes the only type of phototrophy found in polar Octadecabacters (Section 4.3.2). Interestingly, among polar roseobacters, which are not members of the Octadecabacter clade, the genes required for xanthorhodopsin synthesis and aerobic anoxygenic photosynthesis (AAP) are equally abundant (present in 3 out of 5 isolates). *Loktanella vestfoldensis* DSM 16212 and *Sulfitobacter guttiformis* KCTC 32187 are genetically capable of both AAP, and xanthorhodopsin phototrophy. On the other hand, the two polar type strains of *L. fryxellensis* and *L. salsilacus* each only possess genes for one of the two: AAP for the former, and xanthorhodopsin for the latter. Phototrophic functions are absent in *Pseudophaeobacter arcticus*. This seems to indicate that the role phototrophy plays for polar Octadecabacters differs from that of most other polar *Roseobacter* group bacteria. Notably, the different micro-environments in which other polar roseobacters dwell pose other restrictions and boundary conditions, most notably the diverse ecosystems of the Antarctic lakes [341]. Thus, in some environments, the additional cost of AAP may become feasible, whereas the benefit conveyed by rhodopsin phototrophy may be marginalised. Similarly, metabolic networks and general survival strategies vary between polar representatives of different clades, which is not feasible to investigate without more sequenced genomes of polar isolates.

Following this thought, it should be noted that the present comparison has important limitations: first, the number of other polar isolates is too low to deduce definitive trends between them and temperate representatives of their associated genera. Owed to that, it is, secondly, not clear whether different polar species form monophyletic clades

within these genera like the polar Octadecabacters do. The two polar type strains of *L. fryxellensis* and *L. salsilacus*, for example, clearly do not, despite their close relation to each other, and *L. vestfoldensis* DSM 16212 falls into a different sub-clade altogether (Figure 3.2). Therefore, these three isolates do not form a distinct group like the polar Octadecabacters, and may simply be derived from independent allopatric speciation events, showing spatially limited distributions and concomitantly smaller population sizes [342]. Such differences in evolutionary background and general ecology reduce the explanatory power of a comparison between both groups.

Nonetheless, these limitations also underline the Octadecabacter clade's current unique position among the roseobacters: the fact that its polar isolates do indeed form a common sub-clade, which contains comparatively many sequenced representatives, makes this group the most feasible to study adaptations to polar habitats. Furthermore, the two type strains of *O. arcticus* and *O. antarcticus* are stenopsychrophilic, meaning that they require temperatures for growth below ≈16 °C. All other included polar roseobacter isolates are merely eurypsychrophilic, which means they are able to grow above 16 °C, and show higher optimum growth temperatures. Given more polar isolates throughout other *Roseobacter* group clades, it would be interesting to see how stenopsychrophily is distributed as a trait, and how steno- and eurypsychrophilic adaptations differ within this group of organisms.

4.4 GLOBAL DISTRIBUTION OF OCTADECABACTER-RELATED SEQUENCES

While the environment, from which an individual Octadecabacter strain was isolated, may give an indication of its preferred lifestyle, this thesis sought to more closely investigate biome preferences and the global distribution of Octadecabacters by means of metagenomic read mapping. Indeed, looking at different Octadecabacter groups, some habitat preferences became apparent (Section 3.9). For example, the highest frequencies of mapped nucleotides observed for groups I, II, and III occurred at high, intermediate, and low latitudes, respectively. Group I was preferentially detected in waters of the North Atlantic Ocean, and group II showed closer association with coral reefs, algal blooms, and coastal waters in general (Figures 3.26b and 3.27). Within the Octadecabacters, the observations made for group II most closely mirror the habitats with the highest abundance of roseobacters in general [24], whereas the polar group's distribution resembles that of a cosmopolitan roseobacter phylotype found at the poles and the North Sea [22]. Notably, most of the metagenomes used in this analysis were sampled around the North Sea, as well as the Pacific and Atlantic coasts of North America (Figure 3.25). The southern hemisphere, on

the other hand, is represented by less samples, which will influence the perceived distribution patterns.

Regardless, metagenomes world-wide contain sequences which are highly similar to Octadecabacters, particularly to groups I and III (Figure 3.25). Not all of these metagenomes likely contain sequences which are truly derived from Octadecabacters, because certain effects lead to false positives in both the prediction method of mg-RAST and the mapping approach used here. mg-RAST assigns reads to taxa by their highest BLAST-based similarity, which assumes that an aligned hit belongs to a close phylogenetic neighbour of the respective taxon. This assumption is not valid for reads stemming from rare genes [343], and these will therefore lead to false detection events. The mapping approach tried to contain this effect by only reporting alignments with >90 % identity, and consequently did not return hits in all of the candidate metagenomes (Figure 3.27). Nonetheless, some bodily fluid and faeces metagenomes contained sequences with high similarity to Octadecabacters, which led to their detection in these habitats. Although these detection events are very few compared to the predictions made by mg-RAST, this precaution evidently cannot prevent false hits from influencing the perceived habitat preferences of Octadecabacter groups. False detection events may contribute to the observation that polar Octadecabacters are the most abundant and widely spread group, since they possess large shell-genomes (Section 4.1.2), and potentially contain the most rare genes within the clade. The influence of this effect on the observed distribution patterns is currently unclear. An investigation of the global distribution and abundance of (ideally full-length) rRNA sequences related to Octadecabacters could complement the present analysis, and allow for a more wholesome view. However, such an examination is beyond the focus of this thesis.

Despite the noted limitation, observations made here suggest a cosmopolitan distribution of Octadecabacter group I, and correspond well to an already known phylotype, as mentioned above. This could correlate with their high evolvability (Section 4.3.1), in the same way that occupation of various ecological niches in the *Roseobacter* group is attributed to their genomic flexibility [69]. More precisely, the polar Octadecabacters' high mutagenic potential could only partly have resulted from adaptation to their extreme habitat, and may rather constitute a predisposition of this particular sub-clade, which facilitated both their cold-adaptation and cosmopolitan distribution (see Sections 4.3.3 and 4.3.4 for additional details). As a consequence, the genus associated with this group (see Section 4.1.3) may also contain non-polar, or mesophilic species which share some of the polar Octadecabacters' characteristics when compared to other roseobacters, e. g. a larger mobilome. Vollmers already noted an indication for this in an analysis of 16S rRNA gene sequences, where some sequences from non-polar isolates clustered together with those of *O. arcticus* and *O. antarcticus* [61].

There may thus also exist an alternative means of genetic exchange to the one proposed by Vollmers: he argued that the high degree of similarity between both genomes despite their bipolar distribution indicates some form of genetic exchange, for which he proposed deep-sea water currents, which could transport members of both species across tropical regions, which would normally be too warm for both to survive [61]. Stable exchange of genetic material would however also be favourable between polar members of this genus and highly similar, but not stenopsychrophilic species, the latter of which could distribute this material more widely. This would lead to an indirect genetic exchange between both poles via several closely related intermediate organisms, rather than a direct exchange through transport of the stenopsychrophilic species themselves. However, it should also be noted that the pairwise sequence similarity observed between the polar Octadecabacter strains is typical for what would be expected for members of a genus (curve I in Figure 4.7), and may therefore not be particularly special at all. It could simply imply that arctic and Antarctic species divided from their common ancestor a relatively short time ago. High sequence similarity in places such as the rRNA operon can also result from functional restraint due to the low temperature and high stress levels, restricting the number of viable mutations.

Concerning global distribution, the mapping exposed a converse situation for Octadecabacter group III: this group with two mesophilic members was the only one detected in one of the northern-most sampled metagenomes (Figure 3.25), which does not necessarily result from unspecific mapping events. Rather, it could be a further example of polar habitat colonisation, which has occurred in single lineages throughout multiple clades of the *Roseobacter* group (see also Section 4.3.4).

The frequencies, in which nucleotides mapped to different Octadecabacter groups, vary largely between metagenomes (Figure 3.26). They should nonetheless not be interpreted as a direct indicator of the groups' abundances, for the two following reasons: first, the methodical approach taken here underestimates Octadecabacter content in all metagenomes, since rRNA genes and regions with high sequence identity between roseobacters were excluded from the mapping. This is because due to the high pairwise sequence similarity (>96 % for rRNA genes, see Figure 4.8), and the usually short length of NGS reads, sequences mapped to these regions could not meaningfully distinguish between different Octadecabacters, and would thus scramble the results. Their exclusion therefore increases the reliability of group assignment, at the expense of detection sensitivity.

Secondly, the pre-selection step only included metagenomes in which mg-RAST had taxonomically assigned a percentage of reads to the *Octadecabacter* genus (Section 2.12.1). This assignment only uses publicly available genome information, and was hence restricted to the three

species *O. antarcticus*, *O. arcticus*, and *O. temperatus*, as none of the other genomes were public at the time. The metagenomes in the present analysis are therefore inherently biased towards samples, in which polar Octadecabacters and temperate species of group II occur more frequently. The fact that group IV exhibits the lowest frequencies of mapped nucleotides likely results in part from this bias, especially as it contains only one strain. Its distribution and habitat preferences are thus harder to derive from the present selection of metagenomes than those of the other groups. Nonetheless, it was noted that it featured more prominently in a number of estuarine habitats and sediment samples (Section 3.9), which could be indicative of its preferred ecological niches.

The metagenomic read mapping performed here intended to look for global occurrence patterns within the Octadecabacter clade. As mentioned above, an exhaustive analysis including estimations of Octadecabacter abundances, and correlations to specific environmental parameters, was not in the scope of this thesis. Nonetheless, it became obvious that the different sub-groups within this clade show distinct distributions. This observation lends ecological significance to the subgroup definition, which was originally purely based on phylogenetic relations within the clade and on the similarity of several genome characteristics (Section 3.2). A more detailed view of the different Octadecabacter groups' habitat preferences, and how these correlate with the group-specific genome, will be an interesting prospect of future investigations.

5

CONCLUSION

The genus *Octadecabacter* originally consisted of only two described and genome-sequenced members, which were both isolated from sea ice. They were considered as "extreme cases" among the *Roseobacter* group concerning their unusually large number of transposable elements, inferred gene duplications, and genome rearrangements. However, due to the lack of directly comparable genome sequences from closely related strains, it remained unclear to what extent these observations are actual evidence of environmental adaptations in the two strains, and how these polar isolates might differ from temperate strains of the same genus on the genome level.

The present thesis expanded the number of available comparison genomes and re-examined genomic characteristics within this genus, with a specific focus on adaptations to life in polar habitats.

A direct result of this examination is the distinction between the genus *Octadecabacter* and the Octadecabacter clade. The latter consists of nine sequenced isolates likely associated with five different genera. Uniquely among roseobacters, four of these isolates stem from extreme polar habitats. They belong to three distinct species, which form a common phylogenetic sub-clade that arguably represents the genus *Octadecabacter* sensu stricto, as it includes the two species which were the basis of this genus' original definition. In combination, their specific genomic functions and genome characteristics reflect the altered nutrient spectra and stress levels encountered at their respective polar habitat, and are broadly consistent with our current understanding of sea ice ecology, and psychrotrophic adaptations described in other organisms.

Divergence in both gene content and sequence similarity within the Octadecabacter clade is similar to other clades in the *Roseobacter* group, which is likely due to the fact that different species within a clade often originate from different micro-environments.

A high amount of transposable elements and genome reshuffling, originally observed in *O. arcticus* and *O. antarcticus*, turned out to be characteristic for polar Octadecabacters in general. The analyses presented here could further generalise these findings, concluding that functions related to all aspects of evolvability are more abundant in polar than in temperate Octadecabacters, and that this evolvability is an important prerequisite for polar habitat adaptation. Conversely, further comparisons among a larger set of roseobacter genome sequences now illustrated that several other *Roseobacter* group members, most of which do not originate from polar habitats, possess similar, or even

higher numbers of transposable elements and unique genes, displacing polar Octadecabacters from their position as extreme cases. This indicates that genome flexibility, while playing an important role in the polar Octadecabacters' evolutionary history, does not seem to be exclusively associated with cold-adaptation in roseobacters. At the same time, adaptation to cold habitats does not generally seem to require higher potential for evolvability, as polar isolates from other clades in the *Roseobacter* group do not exhibit the same characteristics as polar Octadecabacters when compared to their temperate relatives. Polar Octadecabacters thus demonstrate a unique evolutionary path in their environmental adaptation among the currently sequenced roseobacters. This thesis did its best to outline this path based on all currently available genome information. In doing so, it revealed several targets which merit further investigation. For example, for several genes, which were identified as specifically gained or enriched in polar Octadecabacters, the scientific literature did not allow a reasonable prediction of their roles in environmental adaptation. Conversely, some genes were determined to have multiple possible functions by which they would contribute to the habitat adaptation of polar Octadecabacters. Such examples would be good candidates for future experiments to try and clarify their respective roles, as it would further our understanding of the mechanisms behind specific environmental adaptations. It is also noteworthy that the available polar Octadecabacter strains are good candidates for such studies, as they are readily cultivable in a laboratory setting.

Using the example of the nine currently available Octadecabacter genome sequences, the present work also demonstrated how individual metrics and methods of genus delineation are influenced by particular features of a group of organisms. In the *Roseobacter* group, despite their functional divergence, members of a clade typically display a coherence in genetic equipment which goes beyond this clade's core-genome, influencing the way in which delineation of genera is meaningfully possible. From a combination of different metrics, it was concluded that the Octadecabacter clade currently comprises five distinct genera, one of which is formed by the polar (including both Arctic and Antarctic) isolates. However, comparisons with metagenome data sets indicated that this polar Octadecabacter genus may be cosmopolitan, and therefore also contain species from non-polar habitats, which still remain to be isolated. This could also add a novel perspective to the observation that Arctic and Antarctic isolates show high levels of similarity in sequence and gene content despite their vast geographical separation. The mutual genetic exchange required to maintain this degree of similarity may not result from transport of individuals between poles through cold deep water streams, as an early theory surmised [61]. Rather, frequent genetic exchange may occur between exclusively polar Octadecabacters and their more widely distributed

close relatives, the latter mediating genetic information between communities of both poles.

The present work offered a detailed account of the nature of, and possible mechanisms behind, adaptation processes to extreme habitats in some members of the *Roseobacter* group. As such, it serves as a basis for future analyses of similar cases within other roseobacter clades, and for elucidating and modelling adaptation tactics utilised by bacteria in general. By expanding the comparison data set with additional Octadecabacter clade genomes and other novel roseobacter isolates, the analyses presented here yielded a much improved description of this clade than previous studies could, as these were based on only two Octadecabacter isolates. Of course, it can be assumed that a higher number of closely related comparison genomes will further increase the detail and reliability of such descriptions. Therefore, it will be interesting to see this picture evolve, as further genome sequences of the roseobacter group, and possibly the Octadecabacter clade, will no doubt become available in the near future, either through improved isolation and cultivation techniques, or by metagenomic binning, or single cell sorting and sequencing approaches.

APPENDIX

A.1 SUPPLEMENTARY FILES

The supplementary data include the output files of most computations described in Chapter 2. Most visualisations provided throughout the thesis are derived from data contained in these files. Figure A.1 gives an overview of the folder structure and the contained files.

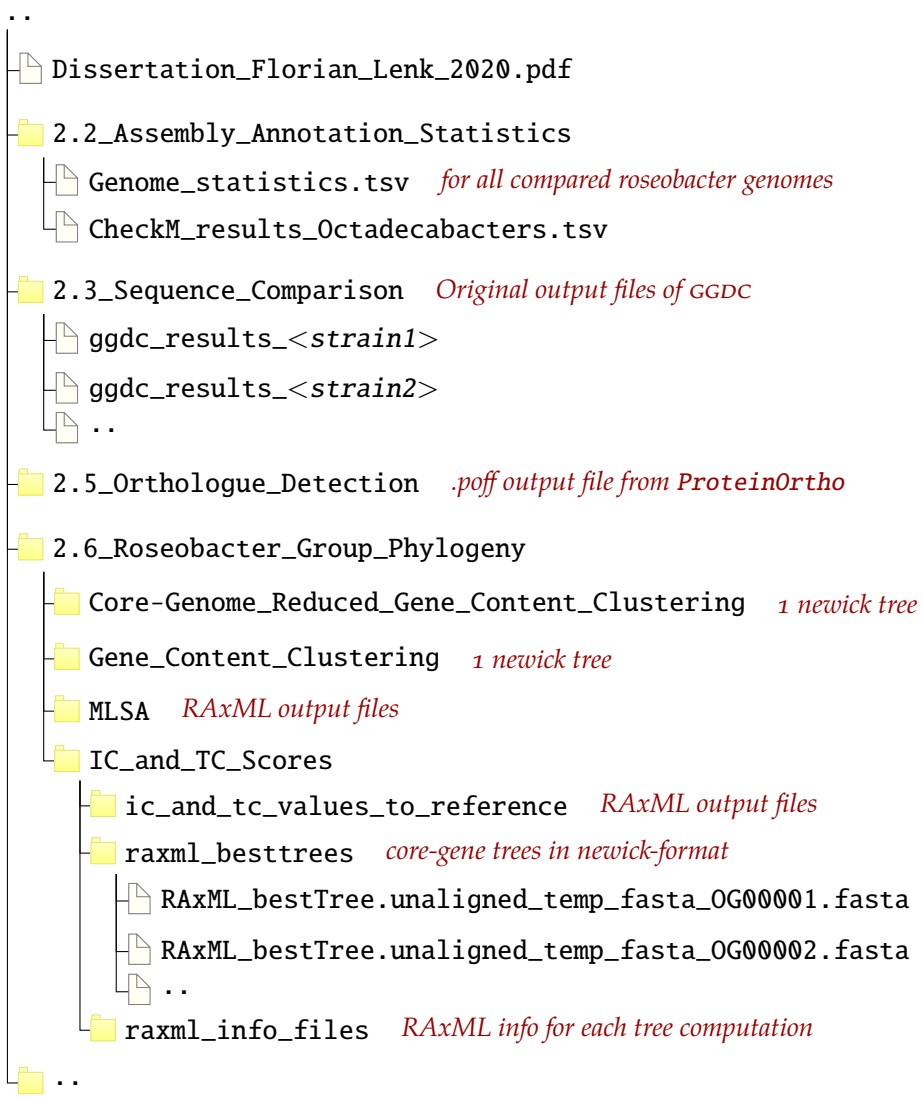


Figure A.1: Supplementary file folder structure and contents (continued on next page).

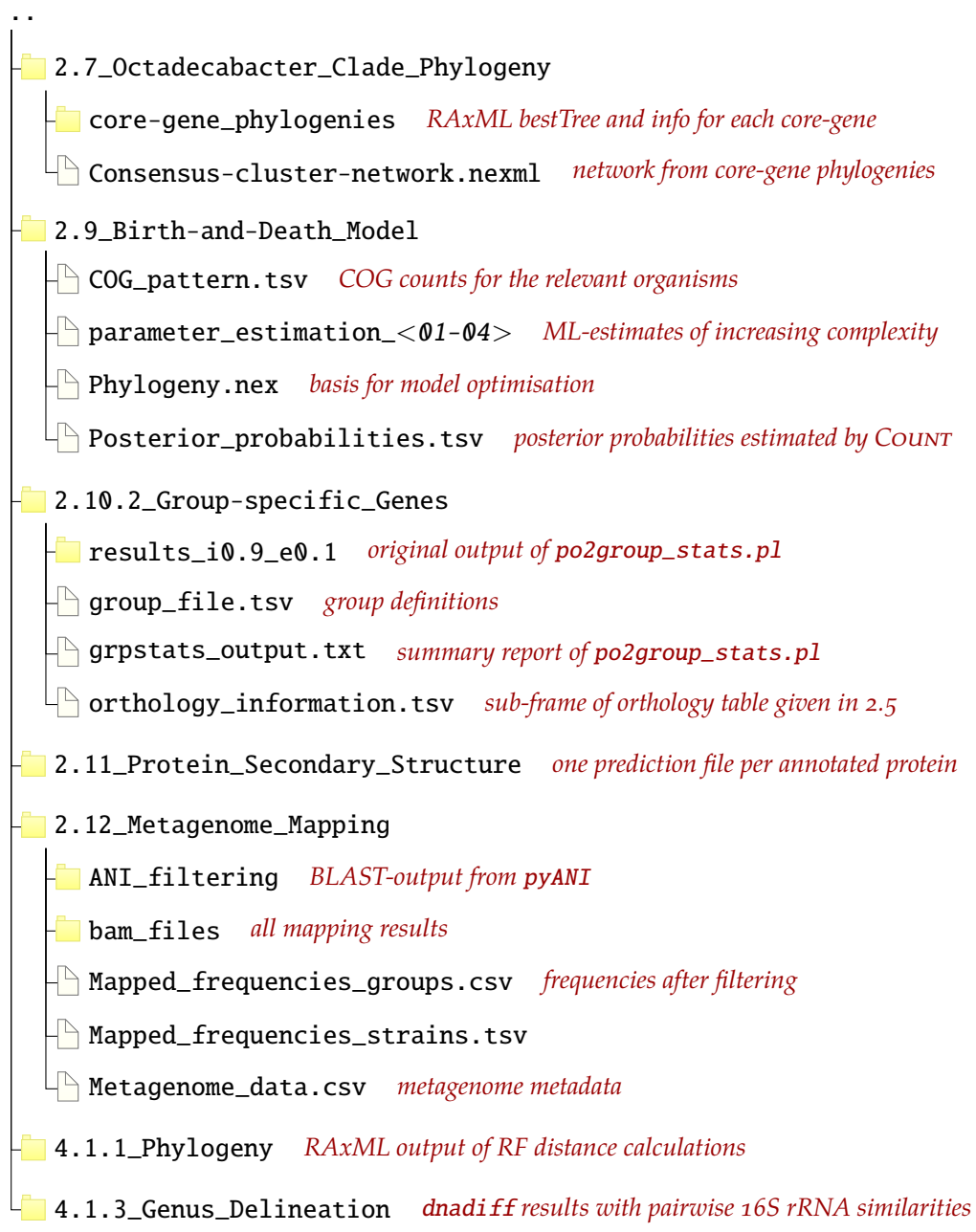


Figure A.1: Supplementary file folder structure and contents (continued)

A.2 ORGANISM METADATA

Table A.1 lists metadata to all strains used for comparative analysis, which were either acquired from the genome data source, or the reference publications given in Table 2.2. Coordinates were either provided directly, or inferred, if the names of specific locations given in the references allowed localisation with reasonable accuracy.

Table A.1: Organism metadata extracted from the database entries, or the respective literature, as available (see Table 2.2). The TEMP. column gives temperature ranges for organism growth, with the optimal growth temperature in grey. HABITAT reflects where the respective organism was isolated and does not necessarily represent its preferred life style.

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Aleiiroseovarius crassostreae	associated host	n.a.	n.d.	LKBA 00000000
Celeribacter baekdonensis B30	sediment	n.a.	n.d.	AMRK 00000000
Celeribacter indicus P73	sediment	63.93E; 27.85S	10–41 (28)	GCA_ 000819565.1
Citreicella sp. 357	sand	42.7748N; 9.1242W	n.d.	AJKJ 00000000
Citreicella sp. SE45	detritus	n.a.	n.d.	ACNW 00000000
Rhodobacterales bacterium HTCC 2255	planktonic	n.a.	n.d.	NZ_AATR 00000000
Phaeobacter sp. LSS9	associated host	n.a.	n.d.	n.a.
Dinoroseobacter shibae DSM 16493	associated patch	n.a.	15–38 (33)	NC_ 009952
Jannaschia aquimarina GSW-M26	planktonic	n.a.	15–37 (30)	NZ_JYFE 00000000
Jannaschia rubra DSM 16279	planktonic	39.4423N; 0.284485W	4-25	n.a.
Jannaschia sp. CCS1	planktonic	38.311N; 123.066W	n.d.	GCA_ 000013565.1
Ketogulonicigenium vulgare WSH-001	n.a.	n.a.	n.d.	NC_ 017384
Ketogulonicigenium vulgare Y25	terrestrial	n.a.	n.d.	GCA_ 000164885
Loktanella atrilutea DSM 29326	planktonic	n.a.	n.d.	NZ_FQUE 00000000
Leisingera aquaemixtae CECT 8399	planktonic	33.251944N; 126.623889E	10–40 (30)	NZ_CYSR 00000000
Leisingera aquimarina DSM 24565	biofilm	44.4123N; 8.9264E	4–37 (20)	NZ_AXBE 00000000

Table A.1: continued

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Leisingera methylohalidivorans DSM 14336	planktonic	38.9N; 77.03W	4–36 (27)	CP 006773.1
Sedimentalea nanhaiensis DSM 24252	sediment	15.55N; 114.49E	4–37 (25)	NZ_AXBG 00000000
Loktanella fryxellensis DSM 16213	microbial mat	77.6099S; 163.1555E	5–25 (25)	NZ_FOCI 00000000
Litoreibacter albidus DSM 26922	associated host	42.902N; 131.75E	4–37 (26)	NZ_FNOI 00000000
Litoreibacter arenae DSM 19593	sand	36.048N; 129.584E	5–35 (30)	NZ_AONI 00000000
Litoreibacter ascidiaceicola DSM 100566	associated host	37.801N; 129.021E	4–34 (25)	NZ_FQUV 00000000
Litoreibacter janthinus DSM 26921	sediment	42.902N; 131.75E	4–37 (26)	NZ_FOYO 00000000
Loktanella koreensis DSM 17925	sand	36.048N; 129.584E	5-30	NZ_FOIZ 00000000
Loktanella litorea DSM 29433	planktonic	n.a.	15–37 (30)	NZ_FOZM 00000000
Loktanella cinnabarina LL-001	sediment	41.17667N; 142.20056E	15–35 (25)	NZ_BATB 00000000
Loktanella hongkongensis DSM 17492	biofilm	22.3N; 114.19E	8–44 (30)	NZ_APGJ 00000000
Loktanella sp. SE62	salt marsh	31.39N; 81.27W	n.d.	n.a.
Loktanella vestfoldensis DSM 16212	microbial mat	68.4564S; 78.1898E	5–37 (20)	NZ_ARNL 00000000
Loktanella vestfoldensis SKA53	planktonic	58.94083N; 11.07944E	n.d.	NZ_AAMS 00000000
Loktanella pyoseonensis DSM 21424	sand	33.328N; 126.842E	4–30 (25)	NZ_FNAT 00000000
Loktanella rosea DSM 29591	sediment	42.908N; 132.349E	4–35 (25)	NZ_FTPR 00000000
Loktanella salsilacus DSM 16199	microbial mat	68.4593S; 78.1884E	5–30 (25)	NZ_FOTF 00000000
Loktanella sediminum DSM 28715	sediment	30.243333N; 128.4475E	10–46 (30)	NZ_FQXB 00000000
Loktanella tamlensis DSM 26879	planktonic	33.5277N; 126.5856E	4–30 (27)	NZ_FOYP 00000000
Marinovum algicola DG 898	associated patch	n.a.	n.d.	GCF_ 001046955

Table A.1: continued

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Maritimibacter alkaliphilus HTCC 2654	n.a.	31.66667N; 64.16667W	16–37 (30)	NZ_AAMT00000000
Parvularcula bermudensis HTCC 2503	planktonic	31.66667N; 64.16667W	10–37 (30)	NZ_AAMU01000000
Maritimibacter sp. HL-12	n.a.	n.a.	n.d.	NZ_FXBQ00000000
Nautella italica DSM 26436	biofilm	44.404N; 8.923E	4–45 (25)	GCA_900113345
Nautella italica R11	associated host	33.99389S; 151.26667E	n.d.	NZ_ABXM00000000
Nereida ignava CECT 5292	planktonic	39.4778N; 0.281W	10-30	NZ_CVPC00000000
Oceanibulbus indolifex HEL-45	planktonic	54.13333N; 7.86667E	n.d.	NZ_ABID00000000
Oceanicola batsensis HTCC 2597	planktonic	31.66667N; 64.16667W	4–40 (30)	NZ_AAMO00000000
Oceanicola granulosus HTCC 2516	planktonic	31.66667N; 64.16667W	4–40 (28)	NZ_AAOT00000000
Oceanicola nanhaiensis DSM 18065	sediment	n.a.	10–37 (28)	NZ_JHZF00000000
Oceanicola sp. HL-35	n.a.	n.a.	n.d.	NZ_JAFT00000000
Oceanicola sp. MCTG1561a	n.a.	56.4397N; 5.5449W	n.d.	NZ_JQMY00000000
Oceanicola sp. S124	planktonic	7.4N; 151.75E	n.d.	GCA_000220565.2
Oceaniovalibus guishaninsula JLT2003	planktonic	24.845N; 121.94E	16–40 (25)	NZ_AMGO00000000
Octadecabacter antarcticus 307	ice	77.883333S; 166.583333E	4-10	GCA_000155735.2
Octadecabacter arcticus DSM 13978	ice	73.016667N; 148.516667E	4-15	GCA_000155675.2
Octadecabacter ascidiaceicola CECT 8868	associated host	34.836N; 128.447E	10–30 (25)	NZ_FXYD00000000
Pseudooctadecabacter jejudonensis CECT 8397	planktonic	33.2525N; 126.6236E	15–35 (30)	n.a.
Octadecabacter sp. E8	associated host	53.703999N; 7.703611E	n.d.	n.a.
Octadecabacter sp. NH9-P7	sediment	53.70466N; 7.70447E	n.d.	n.a.
Octadecabacter temperatus Sb1	planktonic	54.18N; 7.89E	4–25 (20)	GCA_001187845

Table A.1: continued

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Octadecabacter sp. PAMC27224	sediment	74.64611S; 164.22333E	10-25	n.a.
Octadecabacter sp. PAMC27225	sediment	74.64611S; 164.22333E	10-25	n.a.
Pelagibaca bermudensis HTCC 2601	planktonic	31.66667N; 64.16667W	10–40 (30)	NZ_AATQ 00000000
Pseudophaeobacter arcticus DSM 23566	sediment	75.006667N; 169.993611W	0–25 (20)	NZ_AXBF 00000000
Leisingera caerulea DSM 24564	biofilm	44.404N; 8.923E	4–45 (25)	NZ_AXBI 00000000
Leisingera daeponensis DSM 23529	sediment	33.2371N; 126.4248E	4-42	NZ_AXBD 00000000
Phaeobacter gallaeciensis ANG1	associated host	21.269972N; 157.73739W	n.d.	NZ_AFCF 00000000
Phaeobacter gallaeciensis BS107	planktonic	43.391910N; 8.403664W	15–37 (25)	GCA_ 000511385
Phaeobacter gallaeciensis DSM 17395	n.a.	n.a.	n.d.	GCA_ 000154765
Phaeobacter gallaeciensis DSM 26640	planktonic	43.391910N; 8.403664W	15–37 (25)	GCA_ 000511385
Phaeobacter inhibens DSM 16374	planktonic	53.705556N; 7.718722E	4–36 (28)	GCA_ 000473105
Planktomarina temperata RCA23	planktonic	53.7N; 7.71E	10–30 (25)	GCA_ 000738435
Rhodobacteraceae bacterium SB2	planktonic	54.171943N; 7.894196E	n.d.	NZ_LGRT 00000000
Rhodobacterales sp. HTCC 2083	planktonic	44.652137N; 124.070921W	n.d.	NZ_ABXE 00000000
Rhodobacterales sp. HTCC 2150	planktonic	44.651667N; 124.411667W	n.d.	NZ_AAZX 00000000
Rhodobacterales sp. Y4I	planktonic	30.7236N; 81.5111W	n.d.	NZ_ABXF 00000000
Roseobacter denitrificans OCh 114	associated host	35.159315N; 139.616576E	2-30	GCA_ 000014045.1
Roseobacter litoralis Och 149	associated host	n.a.	2-30	GCA_ 000154785.2
Roseobacter sp. AzwK-3b	planktonic	36.84262N; 121.74701W	n.d.	GCA_ 000170875
Roseobacter sp. CCS2	planktonic	38.308333N; 123.3W	n.d.	NZ_AAYB 00000000
Roseobacter sp. GAI101	planktonic	n.a.	n.d.	NZ_ABXS 00000000

Table A.1: continued

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Roseobacter sp. LE17	n.a.	32.84N; 117.2769W	n.d.	n.a.
Roseobacter sp. MED193	planktonic	41.66667N; 2.8E	n.d.	NZ_AANB 00000000
Roseobacter sp. R2A57	planktonic	44.653167N; 124.178167W	n.d.	n.a.
Roseobacter sp. SK209-2-6	n.a.	n.a.	n.d.	NZ_AAYC 00000000
Roseovarius mucosus DSM 17069	associated patch	54.195387N; 7.893448E	15–43 (31)	NZ_AONH 00000000
Roseovarius nubinhibens ISM	planktonic	22.066667N; 74.066667W	10–40 (30)	NZ_AALY 00000000
Roseovarius sp. MCTG1562b	associated patch	56.332304N; 5.748674W	n.d.	NZ_JQLS 00000000
Roseovarius sp. TM1035	associated patch	n.a.	n.d.	NZ_ABCL 00000000
Ruegeria atlantica DSM 5823	sediment	21.41N; 17.893333W	n.d.	NZ_CYPU 00000000
Ruegeria halocynthiae MOLA R1 13b	associated host	48.695278N; 3.143889E	n.d.	NZ_JQEZ 00000000
Ruegeria lacuscaerulensis ITI-1157	n.a.	63.879N; 22.434W	n.d.	NZ_ACNX 00000000
Ruegeria mobilis F1926	planktonic	31.4061S; 91.17758E	n.d.	NZ_CP 015230.1
Ruegeria pomeroyi DSS-3	planktonic	n.a.	10-40	GCA 000011965.2
Ruegeria sp. TM1040	associated patch	n.a.	n.d.	NZ_AAFG 00000000
Ruegeria sp. TW15	associated host	n.a.	10–37 (27)	NZ_AEYW 00000000
Sagittula stellata E-37	planktonic	n.a.	10–41 (30)	NZ_AAYA 00000000
Shimia marina CECT 7688	biofilm	34.772850N; 128.387213E	15–35 (33)	NZ_CYPW 00000000
Shimia sp. SK013	sediment	57.6125N; 8.59E	10–35 (30)	NZ_LAJH 00000000
Sulfitobacter donghicola KCTC 12864	planktonic	37.240459N; 131.866944E	10–31 (25)	NZ_JASF 00000000
Sulfitobacter donghicola JCM 14565	planktonic	37.240459N; 131.866944E	10–31 (25)	NZ_JASF 00000000
Sulfitobacter geojensis MM-124	planktonic	35.0575N; 128.803056E	4–30 (25)	NZ_JASE 00000000

Table A.1: continued

STRAIN	HABITAT	LOC.	TEMP. [°C]	ACCESSION
Sulfitobacter guttiformis KCTC 32187	planktonic	68.521S; 78.270E	4–32 (16)	NZ_JASG00000000
Sulfitobacter mediterraneus KCTC 32188	planktonic	42.516667N; 3.183333E	4–35 (22)	NZ_JASH00000000
Sulfitobacter pontiacus 3SOLIMAR09	planktonic	39.794141N; 2.693348E	n.d.	NZ_AXZR00000000
Sulfitobacter sp. 20 GPM-1509m	n.a.	n.a.	n.d.	NZ_JIBC00000000
Sulfitobacter sp. CB2047	n.a.	n.a.	n.d.	NZ_JPOY00000000
Sulfitobacter sp EE-36	planktonic	n.a.	n.d.	NZ_AALV00000000
Sulfitobacter sp. MCCC 1A00686	associated patch	n.a.	10–37 (28)	NZ_JAMD00000000
Sulfitobacter sp. NAS-14.1	planktonic	33.3N; 74.3W	n.d.	NZ_AALZ00000000
Sulfitobacter sp. NB-68	n.a.	35.658333N; 129.1275E	4–30 (30)	NZ_JASC00000000
Sulfitobacter sp. NB-77	n.a.	35.658333N; 129.1275E	4–30 (25)	NZ_JASD00000000
Sulfitobacter sp. SA11	associated patch	47.727888N; 122.473801W	n.d.	n.a.
Tateyamaria sp. ANG S1	associated host	21.269972N; 157.73739W	n.d.	NZ_JWLL00000000
Thalassiobium sp. R2A62	planktonic	44.65N; 124.167W	n.d.	NZ_ACOA00000000
Thalassobacter stenotrophicus CECT 5294	n.a.	n.a.	13–37 (24)	NZ_CYRX00000000
Wenxinia marina DSM 24838	sediment	22.83021N; 118.088829E	15–42 (36)	NZ_ARAY00000000
Roseovarius atlanticus R12B	planktonic	15.2S; 13.3W	4–45 (27)	NZ_LAXJ00000000
Roseovarius sp. 217	planktonic	50.25N; 4.21667W	n.d.	NZ_AAMV00000000

A.3 *ROSEOBACTER* GROUP PHYLOGENY AND OTHER TREES

The full super-matrix-based phylogeny of 115 roseobacters plus outgroup organism is given in Figure A.2 (see Section 2.6 for method).

Gene-content-based hierarchical clustering (Figure 4.3) was repeated, deleting the core-genes of 22 sub-clades within the phylogeny in all members of that clade prior to computing Jaccard distances (Figure A.3).

In addition, strains were clustered based on the distances (in kilometres) between their isolation coordinates listed in Table A.1 (Figure A.4).

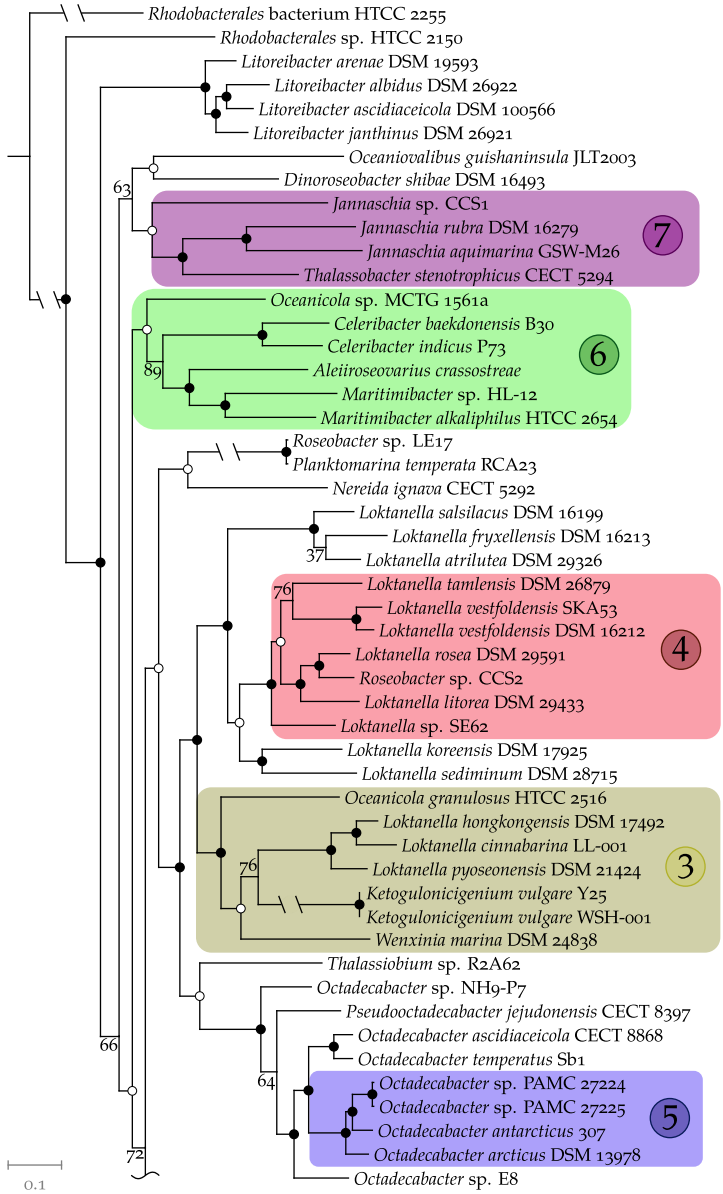


Figure A.2: Super-matrix-based core-genome phylogeny of 115 *Roseobacter* group genomes. Highlighted clades as defined in ref. [59]. Numbers indicate bootstrap support. Open and closed circles represent values of 90–99 and 100, respectively. Scale bar in expected number of substitutions. Long edges were shortened by 50 % (marked by interruptions). Tree continued on next page.

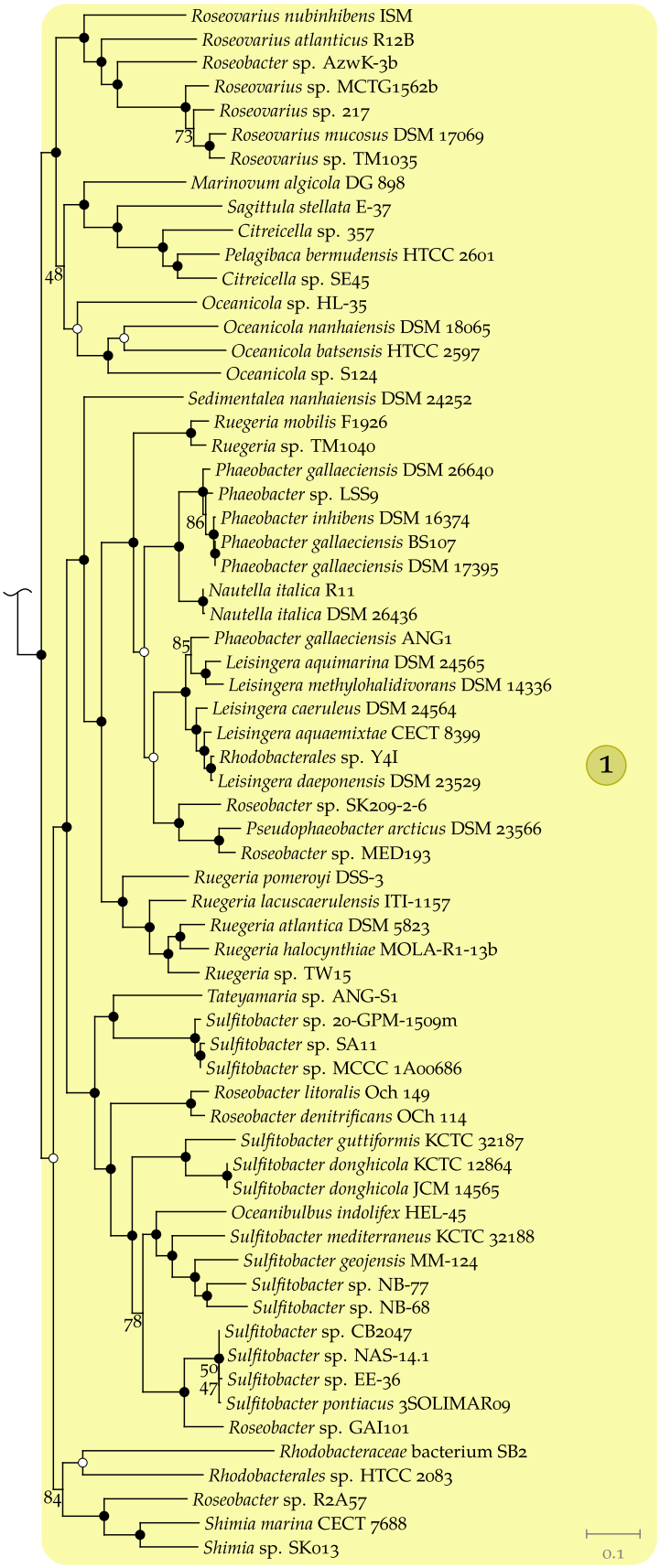


Figure A.2: *Roseobacter* group MLSA phylogeny (continued)

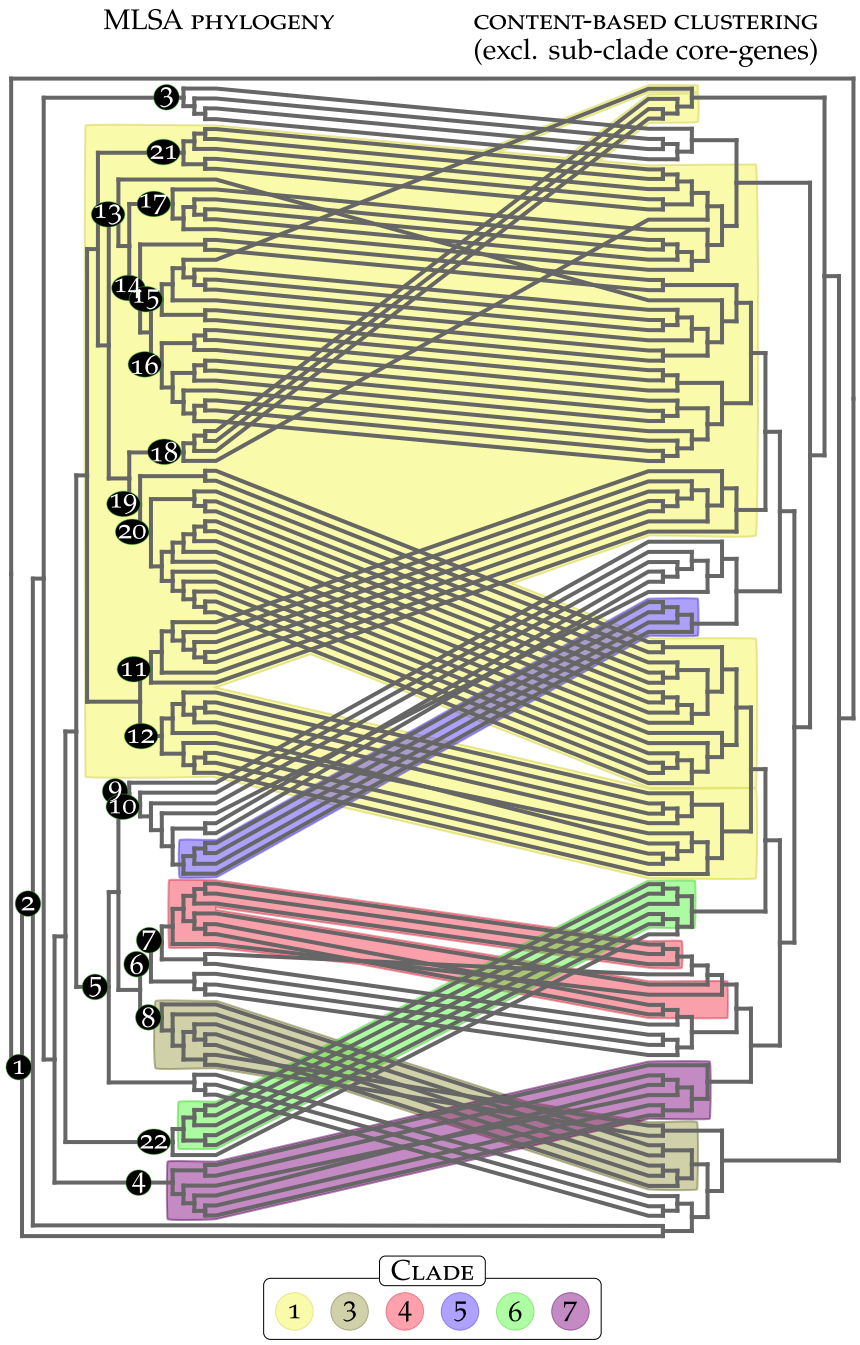


Figure A.3: MLSA phylogeny of 115 roseobacter genomes plus outgroup on the left, as given in Figure A.2, versus hierarchical clustering of the same genomes on the right, based on pairwise Jaccard distances of OG presence and absence. Before clustering, core-genes of sub-clades (labelled 1–22 in the figure) were removed from all members of that clade (that is, they were marked as absent prior to the Jaccard distance calculation) in numerical order. Clades defined in ref. [59] are coloured as in Figure A.2. RF distance between both trees is 0.47, determined using `ete3` [177].

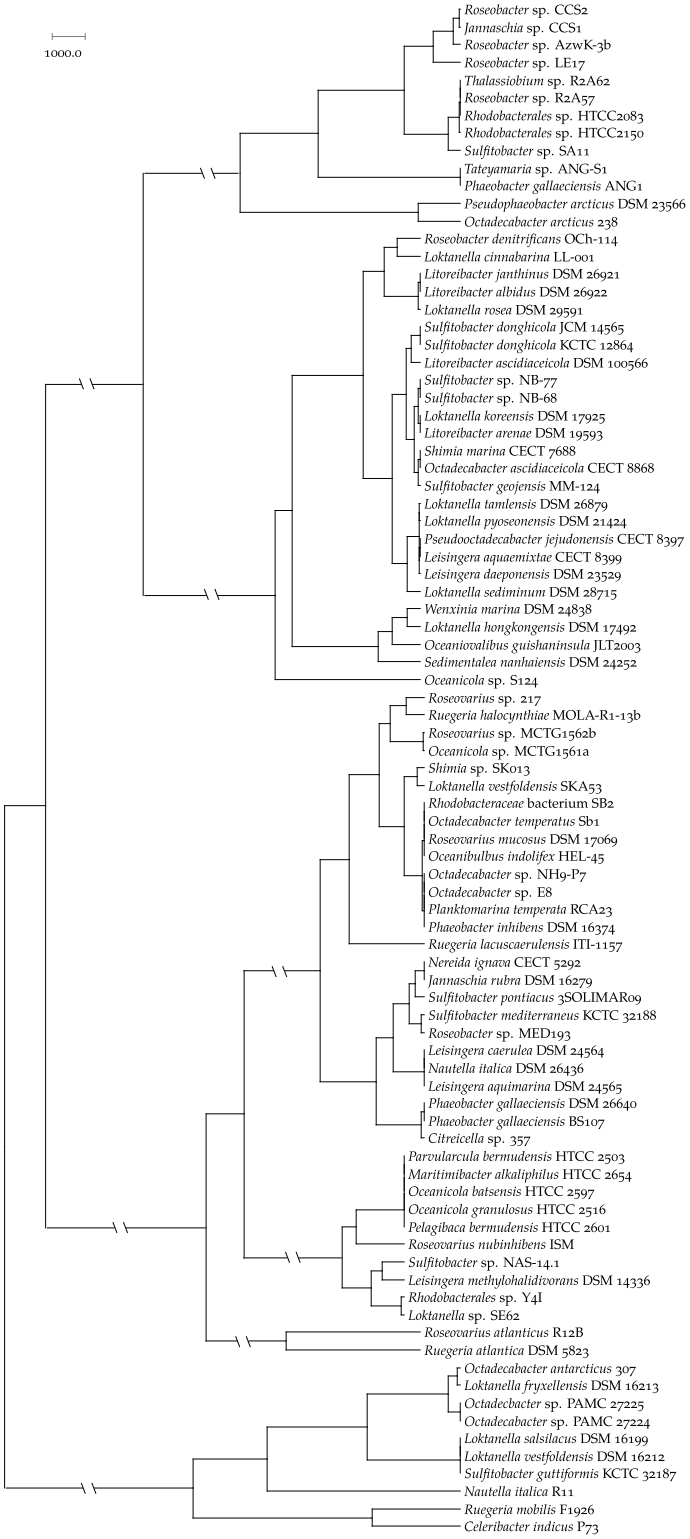


Figure A.4: Hierarchical clustering of genomes based on geographic distance (in kilometres). Only strains with available isolation coordinates (Table A.1) were used in the computation. RF distance to MLSA phylogeny (Figure A.2), gene content clustering, and cloud- and shell-genome clustering (right trees in Figures 4.3 and A.3, respectively) is 0.96, determined using `ete3` [177]. Interrupted edges were shortened by 50 %.

A.4 GROUP-SPECIFIC GENES

Genes specific to each group of Octadecabacter genomes, as defined in Section 2.10.2, are listed in Tables A.2 to A.5 for each group individually.

Table A.2: Genes exclusive to the group formed by *O. arcticus*, *O. antarcticus*, and isolates PAMC 27224 and 27225 (group I, Section 3.2), as determined by `po2group_stats.pl`. For each specific OG, the annotation and LOCUS TAG in the genome of isolate PAMC 27224 is provided for reference.

DESCRIPTION	LOCUS TAG
Site-specific DNA recombinase	octa_00090
Protein of unknown function (DUF3987)	octa_00390
hypothetical protein	octa_02100
hypothetical protein	octa_02490
CBS domain-containing protein	octa_02830
Protein of unknown function (DUF2848)	octa_03810
Uncharacterized conserved protein	octa_04100
RNA polymerase σ^{32} factor	octa_06070
Fe-S cluster biogenesis protein NfuA, 4Fe-4S-binding domain	octa_06080
ATP-binding protein involved in chromosome partitioning	octa_06090
DNA-binding transcriptional regulator, MarR family	octa_06100
Glycine/D-amino acid oxidase (deaminating)	octa_06740
transcriptional regulator, TetR family	octa_09560
aldehyde dehydrogenase (NAD^+)	octa_09610
cysteine desulfurase family protein, VC1184 subfamily	octa_09620
bile acid:Na^+ symporter, BASS family	octa_10390
AAA ATPase domain-containing protein	octa_10590
hypothetical protein (DUF2333)	octa_10600
hypothetical protein	octa_10610
4-α-glucanotransferase	octa_10680
Taurine catabolism dioxygenase TauD, TfdA family	octa_12020
SSU ribosomal protein S21P	octa_12240
TIGR02453 family protein	octa_12280
monovalent cation:H^+ antiporter, CPA1 family	octa_12300
transposase	octa_13300
Uncharacterized membrane protein YdjX, TVP38/TMEM64 family, SNARE-associated domain	octa_13530
selenium-binding protein 1	octa_13670
hypothetical protein	octa_13680
hypothetical protein	octa_14030
β-carotene 15,15'-monooxygenase, Brp/Blh family	octa_14420

Table A.2: continued

DESCRIPTION	LOCUS TAG
lycopene beta-cyclase	octa_14430
phytoene synthase	octa_14440
geranylgeranyl diphosphate synthase, type II	octa_14460
Bacteriorhodopsin	octa_14470
Predicted thiol-disulfide oxidoreductase YuxK, DCC family	octa_14550
hypothetical protein	octa_16050
ATP-binding protein involved in chromosome partitioning	octa_16490
Fe-S cluster biogenesis protein NfuA, 4Fe-4S-binding domain	octa_16500
Acetyltransferase (GNAT) domain-containing protein	octa_19650
Hemolysin-type calcium-binding repeat-containing protein	octa_20610
PilZ domain-containing protein	octa_23420
myo-inositol 2-dehydrogenase / D-chiro-inositol 1-dehydrogenase	octa_25790
3-hydroxyisobutyrate dehydrogenase	octa_25810
pyruvate dehydrogenase E1 component alpha subunit	octa_25900
pyruvate dehydrogenase E1 component beta subunit	octa_25910
2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase)	octa_25920
succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase	octa_25930
hypothetical protein	octa_26240
hypothetical protein	octa_27350
DNA polymerase V	octa_27360
simple sugar transport system ATP-binding protein	octa_28990
Alcohol dehydrogenase, class IV	octa_29950
Acyl-CoA reductase	octa_29960
glutamine synthetase	octa_29970
Predicted N-formylglutamate amidohydrolase	octa_29980
Alcohol dehydrogenase, class IV	octa_30040
carbohydrate ABC transporter substrate-binding protein, CUT1 family (TC 3.A.1.1.-)	octa_30070
carbohydrate ABC transporter membrane protein 2, CUT1 family	octa_30090
carbohydrate ABC transporter membrane protein 1, CUT1 family	octa_30100
glycerol transport system ATP-binding protein	octa_30110
glycerol transport system ATP-binding protein	octa_30120

Table A.2: continued

DESCRIPTION	LOCUS TAG
glycerol transport system permease protein	octa_30150
glycerol transport system permease protein	octa_30160
dihydroxyacetone kinase	octa_30190
multiple sugar transport system ATP-binding protein	octa_30200
carbohydrate ABC transporter ATP-binding protein, CUT1 family	octa_30210
carbohydrate ABC transporter membrane protein 2, CUT1 family	octa_30230
carbohydrate ABC transporter membrane protein 1, CUT1 family	octa_30240
carbohydrate ABC transporter substrate-binding protein, CUT1 family	octa_30250
LacI family transcriptional regulator	octa_30260
pseudoazurin	octa_30490
simple sugar transport system permease protein	octa_30600
simple sugar transport system substrate-binding protein	octa_30610
transcriptional regulator, TetR family	octa_32600
hypothetical protein	octa_32620
transcriptional regulator, RpiR family	octa_32640
Predicted N-formylglutamate amidohydrolase	octa_32650
TRAP-type mannitol/chloroaromatic compound transport system, small permease component	octa_32660
TRAP transporter, DctM subunit	octa_32670
Tat (twin-arginine translocation) pathway signal sequence	octa_32680
glutamine synthetase	octa_32690
Acyl-CoA reductase	octa_32700
4-hydroxybutyrate dehydrogenase	octa_32710
spermidine/putrescine transport system substrate-binding protein	octa_32720
spermidine/putrescine transport system ATP-binding protein	octa_32730
spermidine/putrescine transport system permease protein	octa_32740
spermidine/putrescine transport system permease protein	octa_32750
phosphate transport system permease protein	octa_36790
ribose transport system permease protein	octa_37100
NAD(P)-dependent dehydrogenase, short-chain alcohol dehydrogenase family	octa_37170
C4-dicarboxylate transporter, DctM subunit	octa_37210

Table A.2: continued

DESCRIPTION	LOCUS TAG
Transposase zinc-binding domain-containing protein	octa_38930
Transposase zinc-binding domain-containing protein	octa_38960
Methyltransferase domain-containing protein	octa_40100
methylthioribose-1-phosphate isomerase	octa_41790
L-fuculose-phosphate aldolase	octa_41810
Pyridoxamine 5'-phosphate oxidase	octa_41850
Site-specific DNA recombinase	octa_42090

Table A.3: Genes exclusive to the group formed by *O. ascidiaceicola*, *O. temperatus* (group II, Section 3.2), as determined by `po2group_stats.pl`. For each specific OG, the annotation and LOCUS TAG in the genome of *O. ascidiaceicola* is provided for reference.

DESCRIPTION	LOCUS TAG
hypothetical protein	oasc_00150
hypothetical protein	oasc_00540
hypothetical protein	oasc_00560
Exopolysaccharide synthesis, ExoD	oasc_02050
hypothetical protein	oasc_02120
hypothetical protein	oasc_02260
hypothetical protein	oasc_03570
SnoaL-like domain protein	oasc_04030
Stress response protein NhaX	oasc_04570
hypothetical protein	oasc_04790
Protein MtfA	oasc_05940
hypothetical protein	oasc_07250
Glycosyl transferases group 1	oasc_07940
hypothetical protein	oasc_07950
N-acetylglucosaminyl-diphospho-decaprenol L-rhamnosyltransferase	oasc_07990
hypothetical protein	oasc_08000
hypothetical protein	oasc_08010
hypothetical protein	oasc_08020
Sulfotransferase domain protein	oasc_08030
hypothetical protein	oasc_08070
Bifunctional hemolysin/adenylate cyclase precursor	oasc_08090
MAPEG family protein	oasc_08820
PilZ domain protein	oasc_09680
hypothetical protein	oasc_09750

Table A.3: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	oasc_09770
hypothetical protein	oasc_09940
hypothetical protein	oasc_10130
hypothetical protein	oasc_10370
Toluene efflux pump outer membrane protein TtgF precursor	oasc_10420
hypothetical protein	oasc_10530
hypothetical protein	oasc_11970
Tellurite resistance protein TerB	oasc_12800
hypothetical protein	oasc_13800
putative xanthine dehydrogenase subunit A	oasc_14730
hypothetical protein	oasc_14810
hypothetical protein	oasc_15100
hypothetical protein	oasc_15400
DNA polymerase IV	oasc_15410
hypothetical protein	oasc_15540
hypothetical protein	oasc_15830
hypothetical protein	oasc_15840
hypothetical protein	oasc_16280
Bacterial SH3 domain protein	oasc_16460
hypothetical protein	oasc_17860
hypothetical protein	oasc_18580
hypothetical protein	oasc_19010
Aquaporin Z 2	oasc_19360
hypothetical protein	oasc_20210
Spermidine synthase	oasc_20220
Universal stress protein F	oasc_20280
hypothetical protein	oasc_21230
α-ketoglutaric semialdehyde dehydrogenase	oasc_21950
hypothetical protein	oasc_22120
hypothetical protein	oasc_22410
hypothetical protein	oasc_22910
hypothetical protein	oasc_23140
Demethylmenaquinone methyltransferase	oasc_24630
3-methyl-adenine DNA glycosylase I	oasc_25850
hypothetical protein	oasc_26640
Bifunctional hemolysin/adenylate cyclase precursor	oasc_27230
hypothetical protein	oasc_27250
hypothetical protein	oasc_27410

Table A.3: continued

DESCRIPTION	LOCUS TAG
Sensor protein EvgS precursor	oasc_28280
hypothetical protein	oasc_28300
Poly-beta-1,6-N-acetyl-D-glucosamine synthase	oasc_28310
Phytanoyl-CoA dioxygenase (PhyH)	oasc_30060
hypothetical protein	oasc_30790
hypothetical protein	oasc_32470
Antitoxin ParD4	oasc_32670

Table A.4: Genes exclusive to the group formed by *P. jejudonensis* and isolate NH9-P7 (group III, Section 3.2), as determined by `po2group_stats.pl`. For each specific OG, the annotation and LOCUS TAG in the genome of *P. jejudonensis* is provided for reference.

DESCRIPTION	LOCUS TAG
Protein of unknown function (DUF3833)	OJEJ_00230
Catechol 2,3-dioxygenase	OJEJ_02950
magnesium chelatase accessory protein	OJEJ_03060
magnesium chelatase subunit D	OJEJ_03070
magnesium chelatase subunit I	OJEJ_03080
spheroidene monooxygenase	OJEJ_03090
phytoene synthase	OJEJ_03110
TspO and MBR related proteins	OJEJ_03120
carotenoid 1,2-hydratase	OJEJ_03130
1-hydroxycarotenoid 3,4-desaturase	OJEJ_03140
farnesyl-diphosphate synthase	OJEJ_03150
demethylspheroidene O-methyltransferase	OJEJ_03160
3-hydroxyethyl bacteriochlorophyllide a dehydrogenase	OJEJ_03170
chlorophyllide a reductase subunit X	OJEJ_03180
chlorophyllide a reductase subunit Y	OJEJ_03200
chlorophyllide a reductase subunit Z	OJEJ_03210
PufQ cytochrome subunit	OJEJ_03220
light-harvesting complex 1 beta chain	OJEJ_03230
light-harvesting complex 1 alpha chain	OJEJ_03240
photosynthetic reaction center L subunit	OJEJ_03250
photosynthetic reaction center M subunit	OJEJ_03260
Intrinsic membrane protein PufX	OJEJ_03270
1-deoxy-D-xylulose-5-phosphate synthase	OJEJ_03290
geranylgeranyl reductase	OJEJ_03310

Table A.4: continued

DESCRIPTION	LOCUS TAG
MFS transporter, BCD family, chlorophyll transporter	OJEJ_03320
chlorophyll synthase	OJEJ_03330
transcriptional regulator PpsR	OJEJ_03340
Methanogenic corrinoid protein MtbC1	OJEJ_03350
3-vinyl bacteriochlorophyllide hydratase	OJEJ_03360
ferredoxin protochlorophyllide reductase subunit N	OJEJ_03370
ferredoxin protochlorophyllide reductase subunit B	OJEJ_03380
cobaltochelatase CobN subunit	OJEJ_03390
ferredoxin protochlorophyllide reductase subunit L	OJEJ_03410
Mg-protoporphyrin IX methyltransferase	OJEJ_03420
MFS transporter, BCD family, chlorophyll transporter	OJEJ_03430
photosynthetic reaction center H subunit	OJEJ_03440
PH domain-containing protein	OJEJ_03450
putative photosynthetic complex assembly protein	OJEJ_03460
hypothetical protein	OJEJ_03470
Mg-protoporphyrin IX monomethyl ester (oxidative) cyclase	OJEJ_03480
putative photosynthetic complex assembly protein 2	OJEJ_03490
dimethylglycine dehydrogenase	OJEJ_03590
Threonine/homoserine/homoserine lactone efflux protein	OJEJ_04100
divinylchlorophyllide 8-vinylreductase	OJEJ_04740
CubicO group peptidase, β-lactamase class C family	OJEJ_04920
single-strand binding protein	OJEJ_07520
hypothetical protein	OJEJ_08130
TRAP-type C4-dicarboxylate transport system, substrate-binding protein	OJEJ_08520
Tripartite ATP-independent transporter, DctQ component	OJEJ_08530
Hemolysin-type calcium-binding repeat-containing protein	OJEJ_09550
DNA-binding transcriptional regulator, LysR family	OJEJ_10900
cold shock protein (beta-ribbon, CspA family)	OJEJ_11070
hypothetical protein	OJEJ_12280
cytochrome b561	OJEJ_12290
hypothetical protein	OJEJ_12550
hypothetical protein	OJEJ_13220
hypothetical protein	OJEJ_13230
malonyl-CoA/methylmalonyl-CoA synthetase	OJEJ_13850
protein-tyrosine phosphatase	OJEJ_14330

Table A.4: continued

DESCRIPTION	LOCUS TAG
NAD-dependent deacetylase	OJEJ_14340
divinyl protochlorophyllide a 8-vinyl-reductase	OJEJ_14840
monoamine oxidase	OJEJ_15480
hypothetical protein	OJEJ_16740
hypothetical protein	OJEJ_16750
Predicted Zn-dependent protease, minimal metalloprotease (MMP)-like domain	OJEJ_16950
Long-chain fatty acid transport protein	OJEJ_17880
hypothetical protein	OJEJ_18740
hypothetical protein	OJEJ_21600
hypothetical protein	OJEJ_22280
hypothetical protein	OJEJ_23970
alpha-glucosidase	OJEJ_26210
GDP-mannose 4,6-dehydratase	OJEJ_28540
UPF0271 protein	OJEJ_29170
inhibitor of KinA	OJEJ_29180
diguanylate cyclase (GGDEF) domain-containing protein	OJEJ_30050
Uncharacterized conserved protein YndB, AHSA1/ START domain	OJEJ_30720
hypothetical protein	OJEJ_32880
starvation-inducible DNA-binding protein	OJEJ_33070

Table A.5: Genes exclusive to the Octadecabacter-associated isolate E8 (group IV, Section 3.2), as determined by `po2group_stats.pl`. For each specific OG, the annotation and LOCUS TAG in the genome is provided for reference.

DESCRIPTION	LOCUS TAG
Nucleoside-diphosphate-sugar epimerase	Ga0068414_1011
peptide/nickel transport system ATP-binding protein	Ga0068414_1012
peptide/nickel transport system ATP-binding protein	Ga0068414_1013
Multidrug resistance protein	Ga0068414_1021
transcriptional regulator, LacI family	Ga0068414_1024
Phage terminase large subunit (GpA)	Ga0068414_1025
Phage DNA packaging protein, Nu1 subunit of terminase	Ga0068414_1026
Uncharacterized conserved protein YjdB, contains Ig-like domain	Ga0068414_1031

Table A.5: continued

DESCRIPTION	LOCUS TAG
Phage minor tail protein U	Ga0068414_10310
Phage minor tail protein U	Ga0068414_1032
Prophage minor tail protein Z (GPZ)	Ga0068414_1033
Phage Head-Tail Attachment	Ga0068414_1034
DNA packaging protein FI	Ga0068414_1035
Phage major capsid protein E	Ga0068414_1036
Bacteriophage lambda head decoration protein D	Ga0068414_1037
protein C (EC:3.4.21.69). Serine peptidase. MEROPS family S49	Ga0068414_1038
Uncharacterized conserved protein YjdB, contains Ig-like domain	Ga0068414_1039
hypothetical protein	Ga0068414_10410
peptide/nickel transport system permease protein	Ga0068414_1042
peptide/nickel transport system substrate-binding protein	Ga0068414_1043
oligopeptide transport system substrate-binding protein	Ga0068414_1044
oligopeptide transport system permease protein	Ga0068414_1047
oligopeptide transport system permease protein	Ga0068414_1048
transcriptional regulator, XRE family with cupin sensor	Ga0068414_1049
adenosylhomocysteinase	Ga0068414_10525
hypothetical protein	Ga0068414_1055
C-terminal domain of 1-Cys peroxiredoxin	Ga0068414_1061
NitT/TauT family transport system ATP-binding protein	Ga0068414_10610
NitT/TauT family transport system permease protein	Ga0068414_10611
NitT/TauT family transport system permease protein	Ga0068414_10612
MgsA AAA+ ATPase C terminal	Ga0068414_10613
Nucleoside 2-deoxyribosyltransferase	Ga0068414_10614
regulatory protein, lacI family	Ga0068414_10615
transcriptional regulator, GntR family	Ga0068414_10616
NitT/TauT family transport system permease protein	Ga0068414_10617
NitT/TauT family transport system ATP-binding protein	Ga0068414_10618
NitT/TauT family transport system substrate-binding protein	Ga0068414_10619
hypothetical protein	Ga0068414_1062

Table A.5: continued

DESCRIPTION	LOCUS TAG
fumarylpyruvate hydrolase	Ga0068414_10620
4-hydroxy-4-methyl-2-oxoglutarate aldolase	Ga0068414_10621
3-hydroxyisobutyrate dehydrogenase	Ga0068414_10622
Uncharacterized membrane protein YfcA	Ga0068414_10623
ABC transporter	Ga0068414_10625
hypothetical protein	Ga0068414_10626
chromosome partitioning protein, ParB family	Ga0068414_10627
chromosome partitioning protein	Ga0068414_10628
hypothetical protein	Ga0068414_1063
DDE domain-containing protein	Ga0068414_10631
hypothetical protein	Ga0068414_10632
iron(III) transport system ATP-binding protein	Ga0068414_10633
iron(III) transport system substrate-binding protein	Ga0068414_10634
iron(III) transport system permease protein	Ga0068414_10635
4-nitrophenyl phosphatase	Ga0068414_10636
peptide/nickel transport system permease protein	Ga0068414_10638
Type I phosphodiesterase / nucleotide pyrophosphatase	Ga0068414_10639
ADP-ribose pyrophosphatase	Ga0068414_1064
glycerophosphoryl diester phosphodiesterase	Ga0068414_10640
oligopeptide transport system ATP-binding protein	Ga0068414_10641
peptide/nickel transport system substrate-binding protein	Ga0068414_10643
catalase	Ga0068414_1065
hypothetical protein	Ga0068414_10654
hypothetical protein	Ga0068414_10655
Transposase InsO and inactivated derivatives	Ga0068414_10656
GntR family transcriptional regulator	Ga0068414_10657
basic membrane protein A	Ga0068414_10658
simple sugar transport system ATP-binding protein	Ga0068414_10659
simple sugar transport system permease protein	Ga0068414_10660
nucleoside ABC transporter membrane protein	Ga0068414_10661
Nicotinamidase-related amidase	Ga0068414_10662
ribokinase	Ga0068414_10663
uridine phosphorylase	Ga0068414_10664
Uncharacterized protein, contains SIS (Sugar ISomerase) phosphosugar binding domain	Ga0068414_10665
succinate-semialdehyde dehydrogenase / glutarate-semialdehyde dehydrogenase	Ga0068414_10666
hypothetical protein	Ga0068414_10667

Table A.5: continued

DESCRIPTION	LOCUS TAG
transcriptional regulator, RpiR family	Ga0068414_1068
ABC-type nitrate/sulfonate/bicarbonate transport system, substrate-binding protein	Ga0068414_1069
Helix-turn-helix	Ga0068414_1071
hypothetical protein	Ga0068414_10710
Integrase core domain-containing protein	Ga0068414_10711
hypothetical protein	Ga0068414_10712
Uracil DNA glycosylase superfamily protein	Ga0068414_10713
hypothetical protein	Ga0068414_1072
hypothetical protein	Ga0068414_1073
HYR domain-containing protein	Ga0068414_1074
hypothetical protein	Ga0068414_1075
hypothetical protein	Ga0068414_1076
Putative flagellar system-associated repeat	Ga0068414_1077
Site-specific recombinase XerD	Ga0068414_10771
ParB-like nuclease domain-containing protein	Ga0068414_10772
hypothetical protein	Ga0068414_10773
Site-specific DNA recombinase	Ga0068414_10774
hypothetical protein	Ga0068414_10775
protein of unknown function (DUF4102)	Ga0068414_10776
Integrase core domain-containing protein	Ga0068414_10777
Hemolysin-type calcium-binding repeat-containing protein	Ga0068414_10778
hypothetical protein	Ga0068414_1078
hypothetical protein	Ga0068414_1079
hypothetical protein	Ga0068414_1081
putative efflux protein, MATE family	Ga0068414_10812
hypothetical protein	Ga0068414_10813
cephalosporin-C deacetylase	Ga0068414_10814
transcriptional regulator, LacI family	Ga0068414_10815
FAD dependent oxidoreductase	Ga0068414_10816
multiple sugar transport system substrate-binding protein	Ga0068414_10817
multiple sugar transport system permease protein	Ga0068414_10818
multiple sugar transport system permease protein	Ga0068414_10819
multiple sugar transport system ATP-binding protein	Ga0068414_10820
Phosphoglycerate dehydrogenase	Ga0068414_10821
hypothetical protein	Ga0068414_10825
Signal transduction histidine kinase	Ga0068414_10826

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_10827
hypothetical protein	Ga0068414_10828
Phage integrase family	Ga0068414_10829
hypothetical protein	Ga0068414_1083
hypothetical protein	Ga0068414_10830
hypothetical protein	Ga0068414_10831
hypothetical protein	Ga0068414_10832
Phage integrase family protein	Ga0068414_10833
methyltransferase, FkbM family	Ga0068414_1084
Glycosyl transferase family 2	Ga0068414_1085
PRC-barrel domain-containing protein	Ga0068414_10850
Glycosyltransferase sugar-binding region containing DXD motif-containing protein	Ga0068414_1086
Glycosyltransferase involved in cell wall bisynthesis	Ga0068414_1087
hypothetical protein	Ga0068414_10875
Phage integrase family protein	Ga0068414_10876
hypothetical protein	Ga0068414_10877
hypothetical protein	Ga0068414_10878
Transposase	Ga0068414_10879
hypothetical protein	Ga0068414_1088
putative RNA 2′-phosphotransferase	Ga0068414_10880
AraC-type DNA-binding protein	Ga0068414_10884
outer membrane autotransporter barrel domain-containing protein	Ga0068414_10885
hypothetical protein	Ga0068414_10886
transporter family-2 protein	Ga0068414_10887
Acetyltransferase (GNAT) family protein	Ga0068414_10888
Helix-turn-helix domain-containing protein	Ga0068414_10889
Integrase core domain-containing protein	Ga0068414_1089
hypothetical protein	Ga0068414_10890
hypothetical protein	Ga0068414_10891
protein of unknown function (DUF4422)	Ga0068414_1091
oligopeptide transport system ATP-binding protein	Ga0068414_109100
oligopeptide transport system ATP-binding protein	Ga0068414_109101
peptide/nickel transport system permease protein	Ga0068414_109103
peptide/nickel transport system substrate-binding protein	Ga0068414_109104
Short-chain dehydrogenase	Ga0068414_109106
DNA-binding transcriptional regulator, MarR family	Ga0068414_109107
hypothetical protein	Ga0068414_109114

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_109115
hypothetical protein	Ga0068414_109116
Methyltransferase domain-containing protein	Ga0068414_109117
hypothetical protein	Ga0068414_109118
High-affinity nickel-transport protein	Ga0068414_109122
ABC-type uncharacterized transport system, substrate-binding protein	Ga0068414_109123
hypothetical protein	Ga0068414_109128
TRAP-type C4-dicarboxylate transport system, small permease component	Ga0068414_109130
C4-dicarboxylate transporter, DctM subunit	Ga0068414_109131
hypothetical protein	Ga0068414_109132
amino acid ABC transporter ATP-binding protein, PAAT family	Ga0068414_109133
amino acid ABC transporter membrane protein 2, PAAT family (TC 3.A.1.3.-)	Ga0068414_109134
amino acid ABC transporter membrane protein 1, PAAT family	Ga0068414_109135
amino acid ABC transporter substrate-binding protein, PAAT family	Ga0068414_109136
hypothetical protein	Ga0068414_109142
creatinine amidohydrolase	Ga0068414_109146
Uncharacterized membrane protein	Ga0068414_109148
transcriptional regulator, LacI family	Ga0068414_109149
cytidine deaminase	Ga0068414_109150
5-methylthioadenosine/S-adenosylhomocysteine deaminase	Ga0068414_109151
non-specific ribonucleoside hydrolase	Ga0068414_109152
nucleoside-binding protein	Ga0068414_109153
simple sugar transport system permease protein	Ga0068414_109155
simple sugar transport system permease protein	Ga0068414_109156
guanine deaminase	Ga0068414_109157
UDP-glucose 4-epimerase	Ga0068414_109160
Glycosyltransferase involved in cell wall bisynthesis	Ga0068414_109161
hypothetical protein	Ga0068414_109162
Sulfotransferase family protein	Ga0068414_109163
Polysaccharide pyruvyl transferase	Ga0068414_109165
Glycosyl transferase family 2	Ga0068414_109166
Glycosyltransferase involved in cell wall bisynthesis	Ga0068414_109167
hypothetical protein	Ga0068414_109168

Table A.5: continued

DESCRIPTION	LOCUS TAG
Lipopolysaccharide biosynthesis protein, LPS:glycosyltransferase	Ga0068414_109169
Endonuclease/Exonuclease/phosphatase family protein	Ga0068414_10917
Glycosyltransferase, GT2 family	Ga0068414_1092
TctA family transporter	Ga0068414_10930
putative tricarboxylic transport membrane protein	Ga0068414_10931
Tripartite-type tricarboxylate transporter, receptor component TctC	Ga0068414_10932
hypothetical protein	Ga0068414_10933
LysR family transcriptional regulator, cys regulon transcriptional activator	Ga0068414_10934
phosphate uptake regulator, PhoU	Ga0068414_10939
hypothetical protein	Ga0068414_10951
toxin CcdB	Ga0068414_10952
LysR family transcriptional regulator, glycine cleavage system transcriptional activator	Ga0068414_10953
hypothetical protein	Ga0068414_10954
protein of unknown function (DUF885)	Ga0068414_10955
glyoxylate/hydroxypyruvate reductase A	Ga0068414_10956
peptide/nickel transport system substrate-binding protein	Ga0068414_10957
peptide/nickel transport system permease protein	Ga0068414_10959
Hemolysin-type calcium-binding repeat-containing protein	Ga0068414_1096
peptide/nickel transport system ATP-binding protein	Ga0068414_10960
peptidase T. Metallo peptidase. MEROPS family M20B	Ga0068414_10961
hypothetical protein	Ga0068414_10962
ketopantoate reductase	Ga0068414_10963
Predicted dehydrogenase	Ga0068414_10964
peptide/nickel transport system substrate-binding protein	Ga0068414_10967
peptide/nickel transport system permease protein	Ga0068414_10968
hypothetical protein	Ga0068414_10970
peptide/nickel transport system ATP-binding protein	Ga0068414_10971
peptide/nickel transport system ATP-binding protein	Ga0068414_10972
peptide/nickel transport system permease protein	Ga0068414_10973

Table A.5: continued

DESCRIPTION	LOCUS TAG
peptide/nickel transport system permease protein	Ga0068414_10974
peptide/nickel transport system substrate-binding protein	Ga0068414_10975
Nucleotide-binding universal stress protein, UspA family	Ga0068414_10981
TRAP transporter, DctM subunit	Ga0068414_10982
TRAP-type C4-dicarboxylate transport system, small permease component	Ga0068414_10983
TRAP-type C4-dicarboxylate transport system, substrate-binding protein	Ga0068414_10984
amidohydrolase	Ga0068414_10989
gluconate 2-dehydrogenase gamma chain	Ga0068414_10990
gluconate 2-dehydrogenase alpha chain	Ga0068414_10991
Cytochrome c, mono- and diheme variants	Ga0068414_10992
putative membrane protein	Ga0068414_10993
Uncharacterized membrane protein	Ga0068414_10994
Uncharacterized membrane protein	Ga0068414_10995
cytochrome c oxidase subunit 2	Ga0068414_10996
cytochrome c oxidase subunit I+III	Ga0068414_10997
hypothetical protein	Ga0068414_10998
Glycine/D-amino acid oxidase (deaminating)	Ga0068414_10999
hypothetical protein	Ga0068414_110106
TupA-like ATPgrasp	Ga0068414_110107
Integrase core domain-containing protein	Ga0068414_11011
MJ0042 family finger-like domain-containing protein	Ga0068414_110114
hypothetical protein	Ga0068414_110119
Helix-turn-helix domain-containing protein	Ga0068414_11012
transcriptional regulator, AraC family	Ga0068414_110122
hypothetical protein	Ga0068414_110137
YrhK-like protein	Ga0068414_110139
hypothetical protein	Ga0068414_11014
Opacity protein	Ga0068414_110145
hypothetical protein	Ga0068414_110148
alcohol dehydrogenase	Ga0068414_11015
Ankyrin repeat-containing protein	Ga0068414_110158
hypothetical protein	Ga0068414_11016
hypothetical protein	Ga0068414_11018
transcriptional regulator, LysR family	Ga0068414_11019
hypothetical protein	Ga0068414_110194
Cu^{+}-exporting ATPase	Ga0068414_110205

Table A.5: continued

DESCRIPTION	LOCUS TAG
solute:Na^+ symporter, SSS family	Ga0068414_11022
Protein N-acetyltransferase, RimJ/RimL family	Ga0068414_110220
hypothetical protein	Ga0068414_11023
hypothetical protein	Ga0068414_110236
hypothetical protein	Ga0068414_11024
hypothetical protein	Ga0068414_110244
hypothetical protein	Ga0068414_11026
hypothetical protein	Ga0068414_110272
hypothetical protein	Ga0068414_110275
Predicted dehydrogenase	Ga0068414_11028
aldehyde dehydrogenase	Ga0068414_110289
hypothetical protein	Ga0068414_11029
hypothetical protein	Ga0068414_110291
PAS domain S-box-containing protein	Ga0068414_110292
hypothetical protein	Ga0068414_110294
Lipoprotein-anchoring transpeptidase ErfK/SrfK	Ga0068414_110295
hypothetical protein	Ga0068414_110296
DNA-binding transcriptional regulator, MerR family	Ga0068414_110297
zinc transporter, ZIP family	Ga0068414_110298
hypothetical protein	Ga0068414_110299
hypothetical protein	Ga0068414_1103
hypothetical protein	Ga0068414_11030
Alpha/beta hydrolase family protein	Ga0068414_110300
hypothetical protein	Ga0068414_110301
DNA-binding transcriptional regulator, MarR family	Ga0068414_110302
YHYH protein	Ga0068414_110303
hypothetical protein	Ga0068414_110304
MFS transporter, DHA1 family, bicyclomycin/ chloramphenicol resistance protein	Ga0068414_110305
intein N-terminal splicing region	Ga0068414_110309
hypothetical protein	Ga0068414_11031
hypothetical protein	Ga0068414_110312
iron complex transport system substrate-binding protein	Ga0068414_110313
hypothetical protein	Ga0068414_110314
iron complex transport system substrate-binding protein	Ga0068414_110315
iron complex transport system substrate-binding protein	Ga0068414_110319

Table A.5: continued

DESCRIPTION	LOCUS TAG
iron complex transport system substrate-binding protein	Ga0068414_110320
iron complex transport system permease protein	Ga0068414_110321
iron complex transport system permease protein	Ga0068414_110322
iron complex transport system ATP-binding protein	Ga0068414_110323
Coiled stalk of trimeric autotransporter adhesin	Ga0068414_110324
Invasion protein IalB, involved in pathogenesis	Ga0068414_110325
hypothetical protein	Ga0068414_110326
hypothetical protein	Ga0068414_110327
transcriptional regulator, ArsR family	Ga0068414_110332
hypothetical protein	Ga0068414_110333
uncharacterized protein	Ga0068414_110334
hypothetical protein	Ga0068414_110335
hypothetical protein	Ga0068414_110336
hypothetical protein	Ga0068414_110337
protein of unknown function (DUF2088)	Ga0068414_110338
altronate hydrolase	Ga0068414_110339
L-lactate dehydrogenase	Ga0068414_110341
Tripartite-type tricarboxylate transporter, receptor component TctC	Ga0068414_110343
Tripartite tricarboxylate transporter TctB family protein	Ga0068414_110344
putative tricarboxylic transport membrane protein	Ga0068414_110345
DNA-binding transcriptional regulator, GntR family	Ga0068414_110346
Integrase core domain-containing protein	Ga0068414_110347
hypothetical protein	Ga0068414_110348
aromatic-amino-acid transaminase	Ga0068414_11035
hypothetical protein	Ga0068414_11039
hypothetical protein	Ga0068414_1104
hypothetical protein	Ga0068414_11041
hypothetical protein	Ga0068414_11043
Protein of unknown function (DUF1328)	Ga0068414_11045
hypothetical protein	Ga0068414_1105
transcriptional regulator, LacI family	Ga0068414_11052
hypothetical protein	Ga0068414_1107
hypothetical protein	Ga0068414_1108
YcxB-like protein	Ga0068414_11080
hypothetical protein	Ga0068414_11098
Site-specific recombinase XerD	Ga0068414_1111
type I restriction enzyme M protein	Ga0068414_11110

Table A.5: continued

DESCRIPTION	LOCUS TAG
MFS transporter, SET family, sugar efflux transporter	Ga0068414_111107
Excalibur calcium-binding domain-containing protein	Ga0068414_111110
Ca^{2+}-binding protein, RTX toxin-related	Ga0068414_111118
hypothetical protein	Ga0068414_111126
hypothetical protein	Ga0068414_11113
single-strand binding protein	Ga0068414_111164
Uncharacterized conserved protein YbjQ, UPF0145 family	Ga0068414_111189
Site-specific recombinase XerD	Ga0068414_1112
Tetratricopeptide repeat-containing protein	Ga0068414_111202
hypothetical protein	Ga0068414_111215
hypothetical protein	Ga0068414_111217
Predicted dehydrogenase	Ga0068414_111218
glucose-fructose oxidoreductase	Ga0068414_111219
monosaccharide ABC transporter membrane protein, CUT2 family	Ga0068414_111220
monosaccharide ABC transporter membrane protein, CUT2 family	Ga0068414_111221
monosaccharide ABC transporter ATP-binding protein, CUT2 family	Ga0068414_111222
monosaccharide ABC transporter substrate-binding protein, CUT2 family	Ga0068414_111223
Sugar phosphate isomerase/epimerase	Ga0068414_111224
Predicted dehydrogenase	Ga0068414_111225
regulatory protein, lacI family	Ga0068414_111226
substrate-binding protein domain-containing protein	Ga0068414_111227
Inosine-uridine nucleoside N-ribohydrolase	Ga0068414_111228
putative spermidine/putrescine transport system permease protein	Ga0068414_111229
putative spermidine/putrescine transport system permease protein	Ga0068414_111230
putative spermidine/putrescine transport system substrate-binding protein	Ga0068414_111231
L-fucose isomerase	Ga0068414_111233
GntR family transcriptional regulator	Ga0068414_111234
phenylacetaldehyde dehydrogenase	Ga0068414_111235
tagatose 1,6-diphosphate aldolase	Ga0068414_111236
dihydroxyacetone kinase DhaK subunit	Ga0068414_111237
dihydroxyacetone kinase DhaL subunit	Ga0068414_111238

Table A.5: continued

DESCRIPTION	LOCUS TAG
D-psicose/D-tagatose/L-ribulose 3-epimerase	Ga0068414_111239
acetyl-CoA synthetase	Ga0068414_111240
Rubredoxin-like zinc ribbon domain (DUF35 N)	Ga0068414_111241
acetyl-CoA C-acetyltransferase	Ga0068414_111242
Major Facilitator Superfamily protein	Ga0068414_111243
gamma-glutamyltranspeptidase / glutathione hydrolase	Ga0068414_111244
D-3-phosphoglycerate dehydrogenase	Ga0068414_111245
Hemolysin-type calcium-binding repeat-containing protein	Ga0068414_111247
iron complex transport system ATP-binding protein	Ga0068414_111249
iron complex transport system substrate-binding protein	Ga0068414_111251
(2Fe-2S) ferredoxin	Ga0068414_111252
hypothetical protein	Ga0068414_111254
regulatory protein, luxR family	Ga0068414_111256
hypothetical protein	Ga0068414_111257
hypothetical protein	Ga0068414_111258
hypothetical protein	Ga0068414_111261
hypothetical protein	Ga0068414_111262
hypothetical protein	Ga0068414_111263
hypothetical protein	Ga0068414_111264
Plasmid recombination enzyme	Ga0068414_111265
hypothetical protein	Ga0068414_111266
hypothetical protein	Ga0068414_111267
hypothetical protein	Ga0068414_111268
Signal transduction histidine kinase	Ga0068414_111269
PAS domain S-box-containing protein	Ga0068414_111270
hypothetical protein	Ga0068414_111279
hypothetical protein	Ga0068414_111288
hypothetical protein	Ga0068414_111291
hypothetical protein	Ga0068414_1113
glutathione S-transferase	Ga0068414_111311
hypothetical protein	Ga0068414_111313
DNA-binding transcriptional regulator, IscR family	Ga0068414_111314
Thioredoxin reductase	Ga0068414_111315
FAD/FMN-containing dehydrogenase	Ga0068414_111316
EcoRII C terminal	Ga0068414_111317
putative spermidine/putrescine transport system permease protein	Ga0068414_111319

Table A.5: continued

DESCRIPTION	LOCUS TAG
putative spermidine/putrescine transport system permease protein	Ga0068414_111320
putative spermidine/putrescine transport system substrate-binding protein	Ga0068414_111321
hypothetical protein	Ga0068414_111323
hypothetical protein	Ga0068414_111324
transcriptional regulator, IclR family	Ga0068414_111325
acetolactate synthase-1/2/3 large subunit	Ga0068414_111326
transcriptional regulator, LacI family	Ga0068414_111330
hypothetical protein	Ga0068414_111332
hypothetical protein	Ga0068414_111333
adenosylhomocysteinase	Ga0068414_111334
adenosylhomocysteinase	Ga0068414_111335
hypothetical protein	Ga0068414_1114
hypothetical protein	Ga0068414_1115
Site-specific DNA recombinase	Ga0068414_1116
hypothetical protein	Ga0068414_1117
hypothetical protein	Ga0068414_11177
type I restriction enzyme, R subunit	Ga0068414_1118
type I restriction enzyme, S subunit	Ga0068414_1119
hypothetical protein	Ga0068414_11197
Transposase and inactivated derivatives	Ga0068414_1121
hypothetical protein	Ga0068414_11210
zinc transporter, ZIP family	Ga0068414_112122
hypothetical protein	Ga0068414_112135
DNA-binding transcriptional regulator, MerR family	Ga0068414_112136
intein N-terminal splicing region	Ga0068414_112137
Uncharacterized conserved protein YdeI, YjbR/CyaY-like superfamily, DUF1801 family	Ga0068414_112138
Ubiquinone/menaquinone biosynthesis C-methylase UbiE	Ga0068414_11215
Lrp/AsnC family transcriptional regulator, leucine-responsive regulatory protein	Ga0068414_112153
leucine dehydrogenase	Ga0068414_112154
Uncharacterized conserved protein, DUF1330 family	Ga0068414_112155
hypothetical protein	Ga0068414_112156
NADPH2:quinone reductase	Ga0068414_112157
Uncharacterized conserved protein, DUF1697 family	Ga0068414_112158
hypothetical protein	Ga0068414_112159

Table A.5: continued

DESCRIPTION	LOCUS TAG
Alpha/beta hydrolase family/Bacterial regulatory proteins, luxR family	Ga0068414_112162
hypothetical protein	Ga0068414_112185
transposase	Ga0068414_1122
Activator of Hsp90 ATPase homolog 1-like protein	Ga0068414_112213
Methyladenine glycosylase	Ga0068414_112215
Transposase IS200 like	Ga0068414_112220
Abortive infection C-terminus	Ga0068414_112221
Cell Wall Hydrolase	Ga0068414_112243
Ribosomal protein S18 acetylase RimI	Ga0068414_112251
ribosome-associated heat shock protein Hsp15	Ga0068414_112269
Zn-dependent dipeptidase, dipeptidase homolog	Ga0068414_112275
hypothetical protein	Ga0068414_112299
hypothetical protein	Ga0068414_1123
hypothetical protein	Ga0068414_112327
hypothetical protein	Ga0068414_112332
hypothetical protein	Ga0068414_112336
Uncharacterized membrane protein	Ga0068414_112352
hypothetical protein	Ga0068414_112356
hypothetical protein	Ga0068414_112362
hypothetical protein	Ga0068414_112364
peptide/nickel transport system permease protein	Ga0068414_112370
regulatory protein, luxR family	Ga0068414_112372
Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain	Ga0068414_112383
Site-specific recombinase XerD	Ga0068414_112397
hypothetical protein	Ga0068414_112398
Helix-turn-helix	Ga0068414_112399
Protein of unknown function (DUF3768)	Ga0068414_1124
hypothetical protein	Ga0068414_112400
hypothetical protein	Ga0068414_112401
sporadically distributed protein, TIGR04141 family	Ga0068414_112402
Phage integrase family protein	Ga0068414_112404
hypothetical protein	Ga0068414_112405
hypothetical protein	Ga0068414_112406
hypothetical protein	Ga0068414_112407
hypothetical protein	Ga0068414_112408
replication region DNA-binding N-term	Ga0068414_112409
plasmid mobilization system relaxase	Ga0068414_112410

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_112411
hypothetical protein	Ga0068414_112412
TRAP transporter, 4TM/12TM fusion protein	Ga0068414_112413
Dienelactone hydrolase	Ga0068414_112414
glutathione S-transferase	Ga0068414_112415
peptide/nickel transport system permease protein	Ga0068414_112421
hypothetical protein	Ga0068414_112459
hypothetical protein	Ga0068414_112462
hypothetical protein	Ga0068414_112498
hypothetical protein	Ga0068414_11251
hypothetical protein	Ga0068414_112510
hypothetical protein	Ga0068414_11252
Peptidase M50B-like	Ga0068414_112558
ABC-2 type transport system permease protein	Ga0068414_112561
ABC-2 type transport system permease protein	Ga0068414_112562
HlyD family secretion protein	Ga0068414_112563
transcriptional regulator, TetR family	Ga0068414_112564
regulatory helix-turn-helix protein, lysR family	Ga0068414_112565
hypothetical protein	Ga0068414_112566
Uncharacterized lipoprotein YbaY	Ga0068414_112567
hypothetical protein	Ga0068414_112568
Integrase core domain-containing protein	Ga0068414_112569
Helix-turn-helix domain-containing protein	Ga0068414_112570
hypothetical protein	Ga0068414_112571
hypothetical protein	Ga0068414_112572
transposase	Ga0068414_112573
hypothetical protein	Ga0068414_112575
hypothetical protein	Ga0068414_112576
HupE / UreJ protein	Ga0068414_112577
hypothetical protein	Ga0068414_112578
hypothetical protein	Ga0068414_112579
hypothetical protein	Ga0068414_112580
Transposase DDE domain-containing protein	Ga0068414_112581
hypothetical protein	Ga0068414_112582
Patatin-like phospholipase	Ga0068414_112583
hypothetical protein	Ga0068414_112584
DinB family protein	Ga0068414_112585
Aldo/keto reductase	Ga0068414_112586
luciferase-type oxidoreductase, BA3436 family	Ga0068414_112587

Table A.5: continued

DESCRIPTION	LOCUS TAG
Predicted arabinose efflux permease, MFS family	Ga0068414_112588
DNA-binding transcriptional regulator, LysR family	Ga0068414_112589
hypothetical protein	Ga0068414_112590
Tetratricopeptide repeat-containing protein	Ga0068414_112591
hypothetical protein	Ga0068414_112592
hypothetical protein	Ga0068414_112593
TniQ	Ga0068414_112594
hypothetical protein	Ga0068414_112595
TniQ protein	Ga0068414_112596
AAA domain-containing protein	Ga0068414_112597
putative transposase	Ga0068414_112598
hypothetical protein	Ga0068414_112599
Site-specific recombinase XerD	Ga0068414_1126
hypothetical protein	Ga0068414_112600
HTH-like domain-containing protein	Ga0068414_112603
hypothetical protein	Ga0068414_112604
outer membrane protein, adhesin transport system	Ga0068414_112605
membrane fusion protein, adhesin transport system	Ga0068414_112606
ATP-binding cassette, subfamily C, LapB	Ga0068414_112607
hypothetical protein	Ga0068414_112608
hypothetical protein	Ga0068414_11263
hypothetical protein	Ga0068414_1127
hypothetical protein	Ga0068414_1128
Uncaracterized surface protein containing fasciclin (FAS1) repeats	Ga0068414_11281
hypothetical protein	Ga0068414_11285
hypothetical protein	Ga0068414_1129
hypothetical protein	Ga0068414_1131
HYR domain-containing protein	Ga0068414_11310
hypothetical protein	Ga0068414_1131000
hypothetical protein	Ga0068414_1131005
hypothetical protein	Ga0068414_113101
uncharacterized protein	Ga0068414_1131019
BFD-like [2Fe-2S] binding domain-containing protein	Ga0068414_1131024
hypothetical protein	Ga0068414_1131026
hypothetical protein	Ga0068414_1131027
hypothetical protein	Ga0068414_1131028
solute carrier family 34 (sodium-dependent phosphate cotransporter)	Ga0068414_1131029

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_113103
DNA-binding transcriptional regulator, LysR family	Ga0068414_1131030
transcriptional regulator, IclR family	Ga0068414_1131037
putative tricarboxylic transport membrane protein	Ga0068414_1131038
Tripartite tricarboxylate transporter TctB family protein	Ga0068414_1131039
Crotonobetainyl-CoA:carnitine CoA-transferase CaiB	Ga0068414_1131041
citrate lyase subunit beta / citryl-CoA lyase	Ga0068414_1131042
transcriptional regulator, LacI family	Ga0068414_1131043
putative aldouronate transport system permease protein	Ga0068414_1131044
putative aldouronate transport system permease protein	Ga0068414_1131045
putative aldouronate transport system substrate-binding protein	Ga0068414_1131046
ADP-ribosylglycohydrolase	Ga0068414_1131047
ADP-ribosylglycohydrolase	Ga0068414_1131048
multiple sugar transport system ATP-binding protein	Ga0068414_1131049
Predicted amidohydrolase	Ga0068414_1131060
SIR2-like domain-containing protein	Ga0068414_1131066
hypothetical protein	Ga0068414_1131068
hypothetical protein	Ga0068414_113107
Homeodomain-like domain-containing protein	Ga0068414_1131071
transposase	Ga0068414_1131072
EcoRII C terminal	Ga0068414_1131073
DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains	Ga0068414_1131076
two-component system, NtrC family, sensor kinase	Ga0068414_1131077
Major Facilitator Superfamily protein	Ga0068414_1131078
lysine 2-monooxygenase (EC 1.13.12.2)	Ga0068414_1131079
sulfopropanediol 3-dehydrogenase	Ga0068414_1131082
hypothetical protein	Ga0068414_1131086
Winged helix-turn helix	Ga0068414_11311
hypothetical protein	Ga0068414_1131102
hypothetical protein	Ga0068414_113111
BCCT, betaine/carnitine/choline family transporter	Ga0068414_1131120
hypothetical protein	Ga0068414_1131125
hypothetical protein	Ga0068414_1131153

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_113116
hypothetical protein	Ga0068414_1131164
hypothetical protein	Ga0068414_1131175
Site-specific recombinase XerD	Ga0068414_1131177
hypothetical protein	Ga0068414_1131178
Predicted phosphoesterase, NUDIX family	Ga0068414_1131179
Methyltransferase domain-containing protein	Ga0068414_1131180
hypothetical protein	Ga0068414_1131181
hypothetical protein	Ga0068414_1131183
hypothetical protein	Ga0068414_1131184
hypothetical protein	Ga0068414_1131185
hypothetical protein	Ga0068414_1131186
hypothetical protein	Ga0068414_1131187
hypothetical protein	Ga0068414_1131188
N-acetylmuramoyl-L-alanine amidase	Ga0068414_1131189
Membrane-associated phospholipid phosphatase	Ga0068414_113119
hypothetical protein	Ga0068414_1131190
hypothetical protein	Ga0068414_1131191
hypothetical protein	Ga0068414_1131192
Endonuclease YncB, thermonuclease family	Ga0068414_1131193
Phosphatidylserine/ phosphatidylglycerophosphate/cardiolipin synthase	Ga0068414_1131195
FAD-NAD(P)-binding	Ga0068414_1131196
Uncharacterized conserved protein	Ga0068414_1131197
Transposase InsO and inactivated derivatives	Ga0068414_11312
FMN-dependent oxidoreductase, nitrilotriacetate monooxygenase family	Ga0068414_1131205
Predicted metal-dependent enzyme of the double-stranded beta helix superfamily	Ga0068414_1131210
DNA-binding transcriptional regulator, LysR family	Ga0068414_1131211
His Kinase A (phospho-acceptor) domain-containing protein	Ga0068414_1131216
PAS domain S-box-containing protein	Ga0068414_1131217
Response regulator receiver domain-containing protein	Ga0068414_1131218
RND family efflux transporter, MFP subunit	Ga0068414_1131219
hypothetical protein	Ga0068414_113122
hydrophobe/amphiphile efflux-1 (HAE1) family protein	Ga0068414_1131220

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_1131221
transcriptional regulator, TetR family	Ga0068414_1131231
Predicted flavoprotein CzcO associated with the cation diffusion facilitator CzcD	Ga0068414_1131232
bile acid:Na^+ symporter, BASS family	Ga0068414_1131233
UDP-glucose 4-epimerase	Ga0068414_1131234
Short-chain dehydrogenase	Ga0068414_1131235
hypothetical protein	Ga0068414_113125
hypothetical protein	Ga0068414_1131250
hypothetical protein	Ga0068414_1131259
hypothetical protein	Ga0068414_1131262
sec-independent protein translocase protein TatB	Ga0068414_1131290
Integrase core domain-containing protein	Ga0068414_11313
hypothetical protein	Ga0068414_1131306
TIGR02300 family protein	Ga0068414_113131
dihydroorotase	Ga0068414_1131312
hypothetical protein	Ga0068414_1131315
MFS transporter, DHA1 family, bicyclomycin/ chloramphenicol resistance protein	Ga0068414_1131317
hypothetical protein	Ga0068414_1131320
Uncharacterized membrane protein YoaK, UPF0700 family	Ga0068414_1131366
hypothetical protein	Ga0068414_11314
hypothetical protein	Ga0068414_1131400
hypothetical protein	Ga0068414_1131401
hypothetical protein	Ga0068414_1131402
Sulfotransferase domain-containing protein	Ga0068414_1131404
hypothetical protein	Ga0068414_113142
hypothetical protein	Ga0068414_1131428
hypothetical protein	Ga0068414_113147
hypothetical protein	Ga0068414_1131470
hypothetical protein	Ga0068414_1131471
gamma-glutamyltranspeptidase / glutathione hydrolase	Ga0068414_1131491
zinc transport system substrate-binding protein	Ga0068414_1131494
SIR2-like domain-containing protein	Ga0068414_11315
Ca^{2+}-binding protein, RTX toxin-related	Ga0068414_1131502
hypothetical protein	Ga0068414_1131503
hypothetical protein	Ga0068414_1131504
regulatory protein, luxR family	Ga0068414_1131505

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_1131506
hypothetical protein	Ga0068414_1131508
FAD binding domain-containing protein	Ga0068414_1131509
Cytochrome C oxidase, cbb3-type, subunit III	Ga0068414_1131510
transcriptional regulator, TetR family	Ga0068414_1131511
polar amino acid transport system substrate-binding protein	Ga0068414_1131513
polar amino acid transport system permease protein	Ga0068414_1131514
polar amino acid transport system ATP-binding protein	Ga0068414_1131515
methionine-gamma-lyase	Ga0068414_1131516
hypothetical protein	Ga0068414_1131521
Phage integrase family	Ga0068414_1131522
hypothetical protein	Ga0068414_1131561
DNA-binding protein HU-alpha	Ga0068414_1131563
hypothetical protein	Ga0068414_1131593
Hemolysin-type calcium-binding repeat-containing protein	Ga0068414_1131594
hypothetical protein	Ga0068414_1131595
LysR family transcriptional regulator, glycine cleavage system transcriptional activator	Ga0068414_1131596
hypothetical protein	Ga0068414_1131598
ferredoxin, 2Fe-2S	Ga0068414_1131599
trk system potassium uptake protein TrkH	Ga0068414_11316
Predicted metal-dependent hydrolase, TIM-barrel fold	Ga0068414_1131601
transcriptional regulator, LysR family	Ga0068414_1131603
Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family	Ga0068414_1131605
D-arabinitol 4-dehydrogenase	Ga0068414_113162
transcriptional regulator, AraC family	Ga0068414_113163
Glycosyl hydrolase 108	Ga0068414_1131642
hypothetical protein	Ga0068414_1131666
hypothetical protein	Ga0068414_1131684
trk system potassium uptake protein TrkA	Ga0068414_11317
hypothetical protein	Ga0068414_1131703
His Kinase A (phospho-acceptor) domain-containing protein	Ga0068414_1131704
Glycosyltransferase involved in cell wall bisynthesis	Ga0068414_1131715
hypothetical protein	Ga0068414_1131720

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_1131721
hypothetical protein	Ga0068414_1131729
NitT/TauT family transport system substrate-binding protein	Ga0068414_1131733
transporter, NhaC family	Ga0068414_1131734
Metallopeptidase family M24	Ga0068414_1131735
hypothetical protein	Ga0068414_1131736
ParD-like antitoxin of type II toxin-antitoxin system	Ga0068414_1131737
lactaldehyde dehydrogenase / glycolaldehyde dehydrogenase	Ga0068414_1131746
hypothetical protein	Ga0068414_1131747
hypothetical protein	Ga0068414_1131748
Uncharacterized protein YcnI	Ga0068414_1131749
hypothetical protein	Ga0068414_1131750
hypothetical protein	Ga0068414_1131751
hypothetical protein	Ga0068414_1131752
hypothetical protein	Ga0068414_1131754
Integrase core domain-containing protein	Ga0068414_1131759
hypothetical protein	Ga0068414_1131760
transcriptional regulator, XRE family with cupin sensor	Ga0068414_1131761
peptide/nickel transport system substrate-binding protein	Ga0068414_1131762
peptide/nickel transport system permease protein	Ga0068414_1131764
ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	Ga0068414_1131765
peptide/nickel transport system ATP-binding protein	Ga0068414_1131766
GAF domain-containing protein	Ga0068414_1131767
N-methylhydantoinase A	Ga0068414_1131768
N-methylhydantoinase B	Ga0068414_1131769
Xaa-Pro aminopeptidase	Ga0068414_1131770
N-methylhydantoinase B	Ga0068414_1131771
hypothetical protein	Ga0068414_1131772
Integrase core domain-containing protein	Ga0068414_1131773
hypothetical protein	Ga0068414_1131774
transposase	Ga0068414_1131775
raffinose/stachyose/melibiose transport system permease protein	Ga0068414_1131777
raffinose/stachyose/melibiose transport system permease protein	Ga0068414_1131778

Table A.5: continued

DESCRIPTION	LOCUS TAG
ADP-ribosylglycohydrolase	Ga0068414_1131779
transcriptional regulator, LacI family	Ga0068414_1131781
raffinose/stachyose/melibiose transport system substrate-binding protein	Ga0068414_1131782
glucosamine–fructose-6-phosphate aminotransferase (isomerizing)	Ga0068414_1131783
secondary thiamine-phosphate synthase enzyme	Ga0068414_1131784
Sugar or nucleoside kinase, ribokinase family	Ga0068414_1131785
purine nucleosidase	Ga0068414_1131786
ADP-ribosylglycohydrolase	Ga0068414_1131787
uridine phosphorylase	Ga0068414_1131788
hypothetical protein	Ga0068414_1131789
HAD-superfamily class IIA hydrolase, TIGR01459	Ga0068414_1131791
CDP-alcohol phosphatidyltransferase	Ga0068414_1131793
L-glutamine/L-glutamate/L-aspartate/ L-asparagine ABC transporter membrane protein	Ga0068414_113194
hypothetical protein	Ga0068414_1132
Sulfotransferase family protein	Ga0068414_113208
hypothetical protein	Ga0068414_113212
Hint domain-containing protein	Ga0068414_113225
hypothetical protein	Ga0068414_113228
ATPase	Ga0068414_113234
Na^+/H^+-dicarboxylate symporter	Ga0068414_113244
hypothetical protein	Ga0068414_113249
N-acylneuraminate cytidylyltransferase	Ga0068414_113250
hypothetical protein	Ga0068414_113257
Sulfotransferase family protein	Ga0068414_113260
serine/threonine protein phosphatase 1	Ga0068414_11328
hypothetical protein	Ga0068414_11329
Transposase	Ga0068414_1133
ATPase family associated with various cellular activities (AAA)	Ga0068414_11330
LPXTG-motif cell wall anchor domain-containing protein	Ga0068414_113313
Glutathione-dependent formaldehyde-activating enzyme	Ga0068414_113327
LPS sulfotransferase NodH	Ga0068414_113365
hypothetical protein	Ga0068414_113376
hypothetical protein	Ga0068414_113377
transposase	Ga0068414_1134

Table A.5: continued

DESCRIPTION	LOCUS TAG
hypothetical protein	Ga0068414_113428
hypothetical protein	Ga0068414_113429
hypothetical protein	Ga0068414_113430
hypothetical protein	Ga0068414_113442
hypothetical protein	Ga0068414_113470
quaternary ammonium compound-resistance protein SugE	Ga0068414_113472
hypothetical protein	Ga0068414_113478
nucleoside ABC transporter membrane protein	Ga0068414_113479
simple sugar transport system permease protein	Ga0068414_113480
basic membrane protein A	Ga0068414_113482
Creatinine amidohydrolase	Ga0068414_113483
Phage integrase family protein	Ga0068414_113484
hypothetical protein	Ga0068414_113485
hypothetical protein	Ga0068414_113487
hypothetical protein	Ga0068414_113488
Type VI secretion system VasI, EvfG, VC A0118	Ga0068414_113489
hypothetical protein	Ga0068414_113490
Homeodomain-like domain-containing protein	Ga0068414_1135
hypothetical protein	Ga0068414_113505
extracellular solute-binding protein, family 3	Ga0068414_11357
hypothetical protein	Ga0068414_113579
hypothetical protein	Ga0068414_113580
Protein of unknown function DUF45	Ga0068414_1136
PAS domain-containing protein	Ga0068414_113626
AraC-type DNA-binding protein	Ga0068414_113641
hypothetical protein	Ga0068414_113642
hypothetical protein	Ga0068414_113643
hypothetical protein	Ga0068414_113644
hypothetical protein	Ga0068414_113692
hook-length control protein FliK	Ga0068414_113695
hypothetical protein	Ga0068414_1137
LysR family transcriptional regulator, glycine cleavage system transcriptional activator	Ga0068414_113702
hypothetical protein	Ga0068414_113709
2,4-dienoyl-CoA reductase	Ga0068414_113716
hypothetical protein	Ga0068414_113717
Phage integrase family protein	Ga0068414_113718
methionine aminopeptidase, type I (EC 3.4.11.18)	Ga0068414_113719

Table A.5: continued

DESCRIPTION	LOCUS TAG
FCD domain-containing protein	Ga0068414_113720
Transposase	Ga0068414_113721
His Kinase A (phospho-acceptor) domain-containing protein	Ga0068414_113722
hypothetical protein	Ga0068414_113723
transcriptional regulator, AlpA family	Ga0068414_113724
hypothetical protein	Ga0068414_113725
hypothetical protein	Ga0068414_113726
protein of unknown function (DUF4102)	Ga0068414_113727
mannosyl-3-phosphoglycerate phosphatase	Ga0068414_113728
glucosyl-3-phosphoglycerate synthase	Ga0068414_113729
sucrose phosphorylase	Ga0068414_113730
hypothetical protein	Ga0068414_113745
MarR family protein	Ga0068414_113747
hypothetical protein	Ga0068414_113776
hypothetical protein	Ga0068414_1138
hypothetical protein	Ga0068414_11385
hypothetical protein	Ga0068414_113853
AAA domain-containing protein	Ga0068414_113862
hypothetical protein	Ga0068414_113863
hypothetical protein	Ga0068414_113864
Opacity protein	Ga0068414_113865
hypothetical protein	Ga0068414_113867
peptide/nickel transport system permease protein	Ga0068414_113871
allantoate deiminase	Ga0068414_113875
HTH-like domain-containing protein	Ga0068414_113877
PRC-barrel domain-containing protein	Ga0068414_113880
putative Mg^{2+} transporter-C (MgtC) family protein	Ga0068414_113881
PAS domain S-box-containing protein	Ga0068414_113882
hypothetical protein	Ga0068414_113883
Signal transduction histidine kinase	Ga0068414_113884
hypothetical protein	Ga0068414_113885
hypothetical protein	Ga0068414_113886
Cd^{2+}/Zn^{2+}-exporting ATPase	Ga0068414_113888
hypothetical protein	Ga0068414_113889
L,D-transpeptidase catalytic domain	Ga0068414_113890
transcriptional regulator, MerR family	Ga0068414_113891
NADPH:quinone reductase	Ga0068414_113897
hypothetical protein	Ga0068414_1139

Table A.5: continued

DESCRIPTION	LOCUS TAG
oligopeptide transport system ATP-binding protein	Ga0068414_113919
hypothetical protein	Ga0068414_113935
hypothetical protein	Ga0068414_113953
Glycosyltransferase family 92	Ga0068414_113963
capsular polysaccharide export protein	Ga0068414_113964
hypothetical protein	Ga0068414_11398
hypothetical protein	Ga0068414_113999

A.5 XANTHORHODOPSINS IN *ROSEOBACTER* GROUP BACTERIA

Figure A.5 visualises the isolation coordinates from Table A.1 on the world map, and xanthorhodopsin-bearing strains are highlighted in red. In addition, the amino acid sequences of roseobacter-family xanthorhodopsins were aligned (via `MUSCLE`) with those of select other species, spanning both subgroups of the xanthorhodopsins. A ML phylogeny was reconstructed using `RAxML`'s `HPC` implementation, with the same configuration as given in Section 2.6. Branch support was estimated through 100 bootstrap repetitions.

Figure A.6 presents the final tree, with both xanthorhodopsin subgroups highlighted in different colours. Xanthorhodopsins of roseobacter strains which had not previously been phylogenetically analysed are coloured red. *Roseobacter* group xanthorhodopsins form one highly-supported cluster in close relation to the second subgroup, which was found to be most abundant in cold and predominantly saline habitats [77].

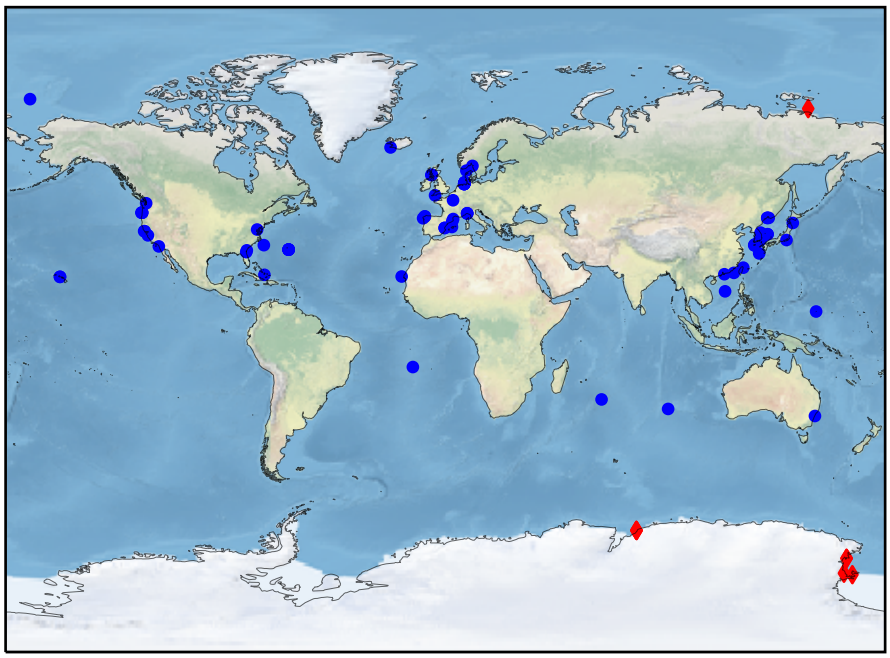


Figure A.5: Isolation coordinates from Table A.1 plotted on the world map. Red diamonds represent strains possessing a xanthorhodopsin.

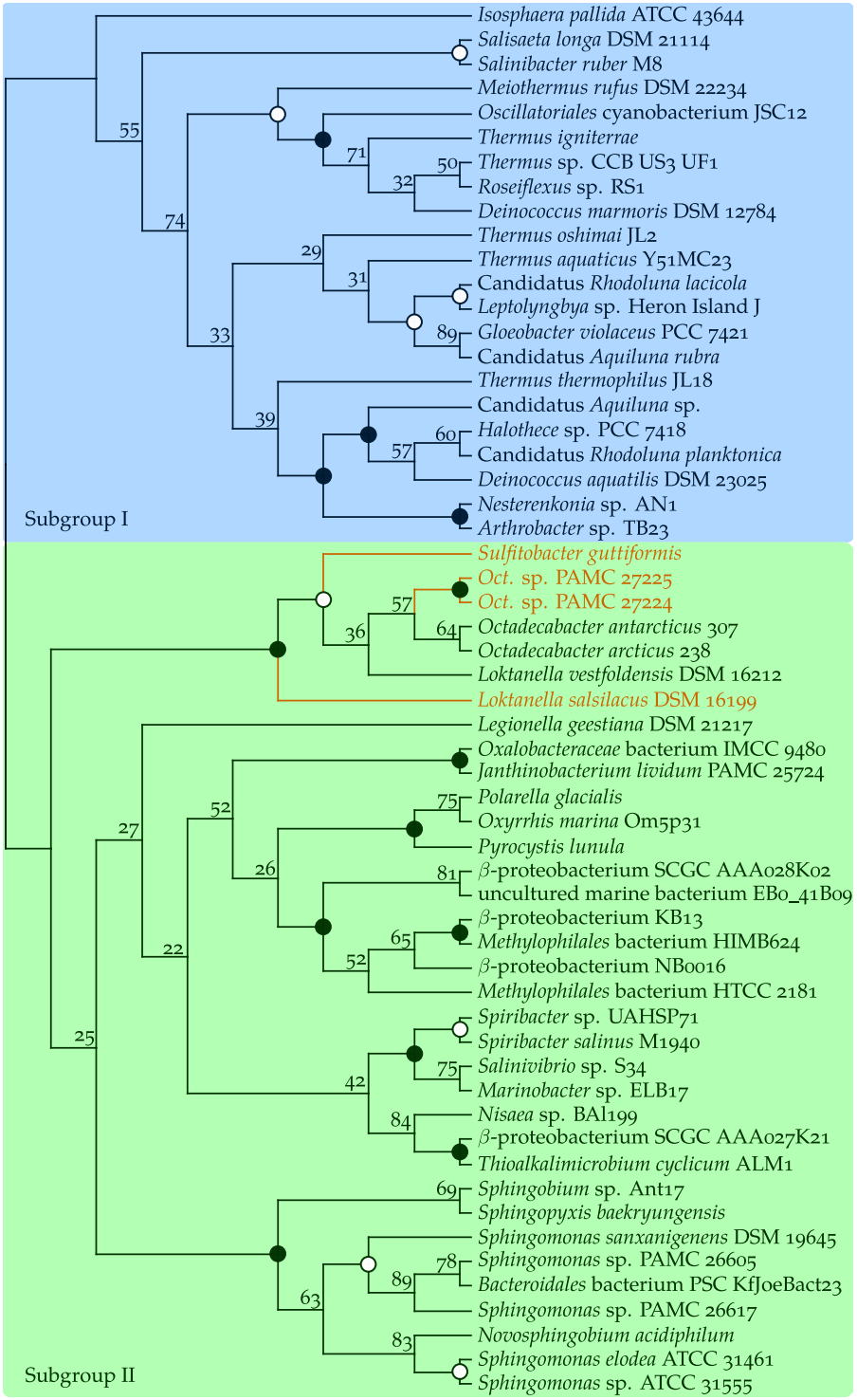
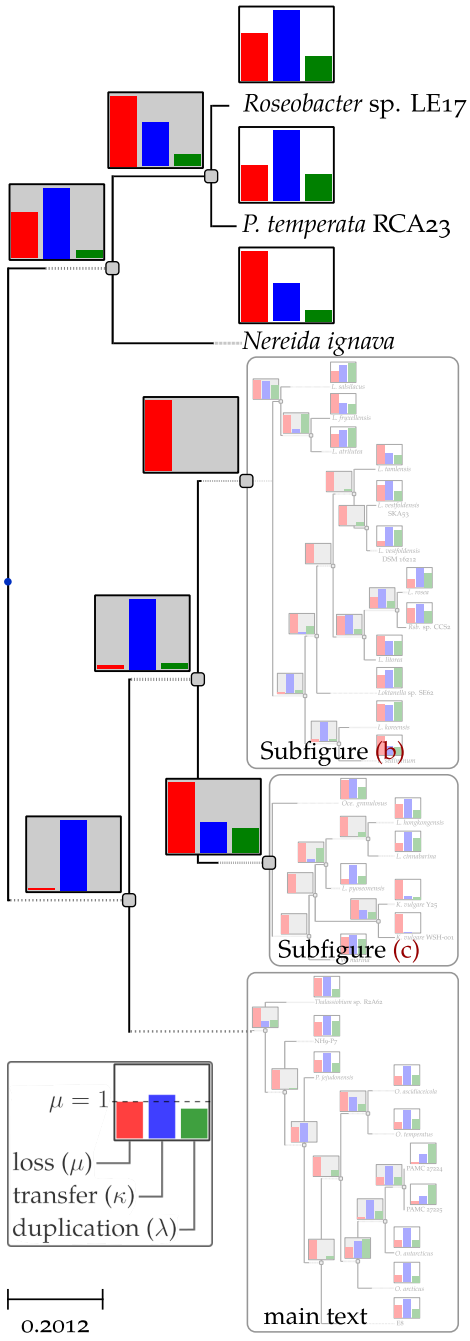


Figure A.6: Phylogenetic placement of *Roseobacter* group xanthorhodopsins. Subgroups as defined in [77]. Numbers indicate bootstrap support, with hollow and filled circles representing values of 90–99, and 100, respectively. Highlighted roseobacter strains' xanthorhodopsins have not previously been classified.

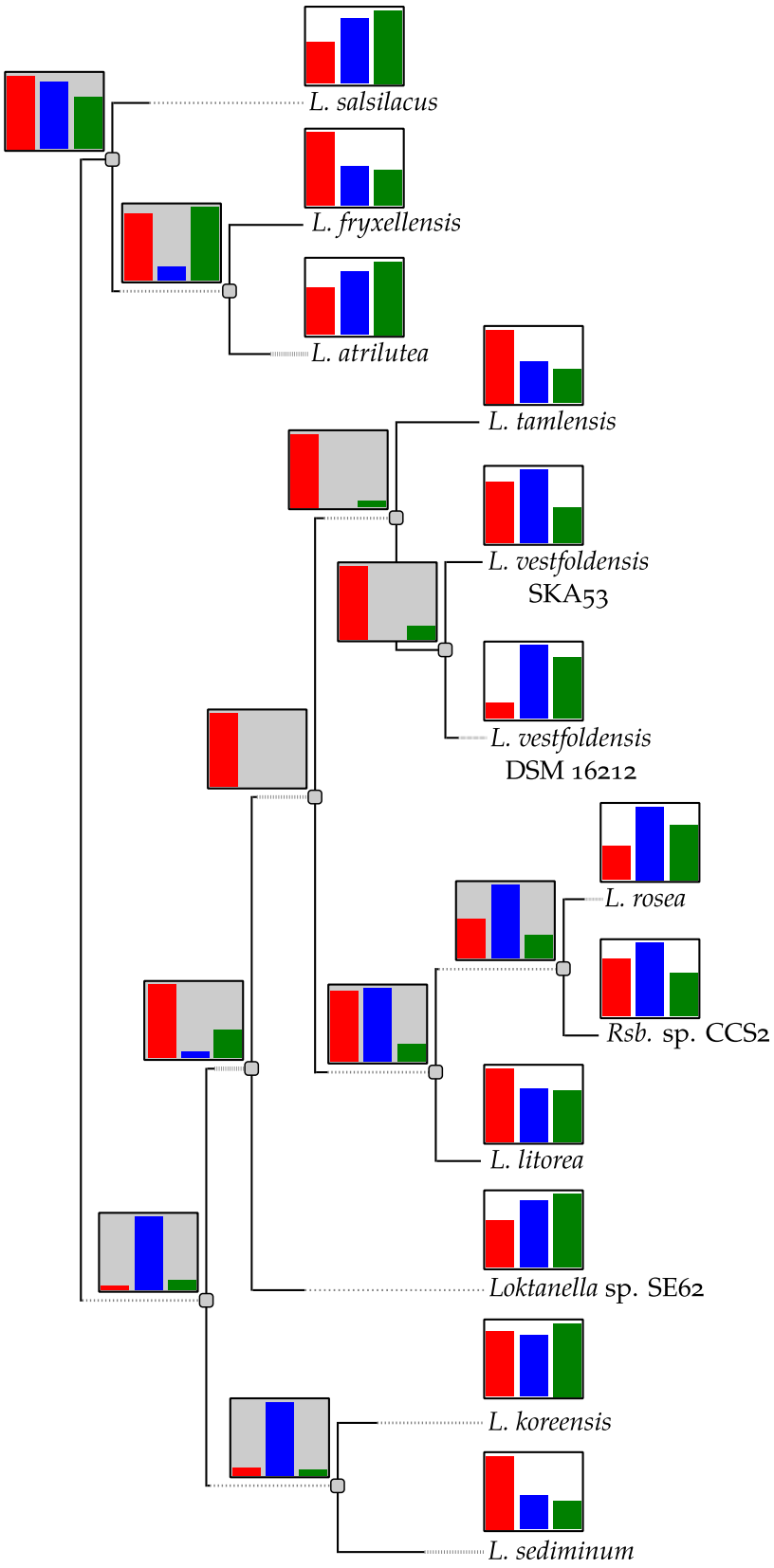
A.6 PHYLOGENETIC BIRTH-AND-DEATH MODEL

Estimated parameter values of the birth-and-death model are visualised for the Octadecabacters in Figure 3.19. Visualisations for the other included strains are given in Figure A.7. Posterior probabilities of COG changes at nodes other than the polar Octadecabacter LCA are provided in Tables A.6 to A.11 (see Figure A.8). They were used to identify changes common to all polar *Roseobacter* group isolates (Section 3.5).



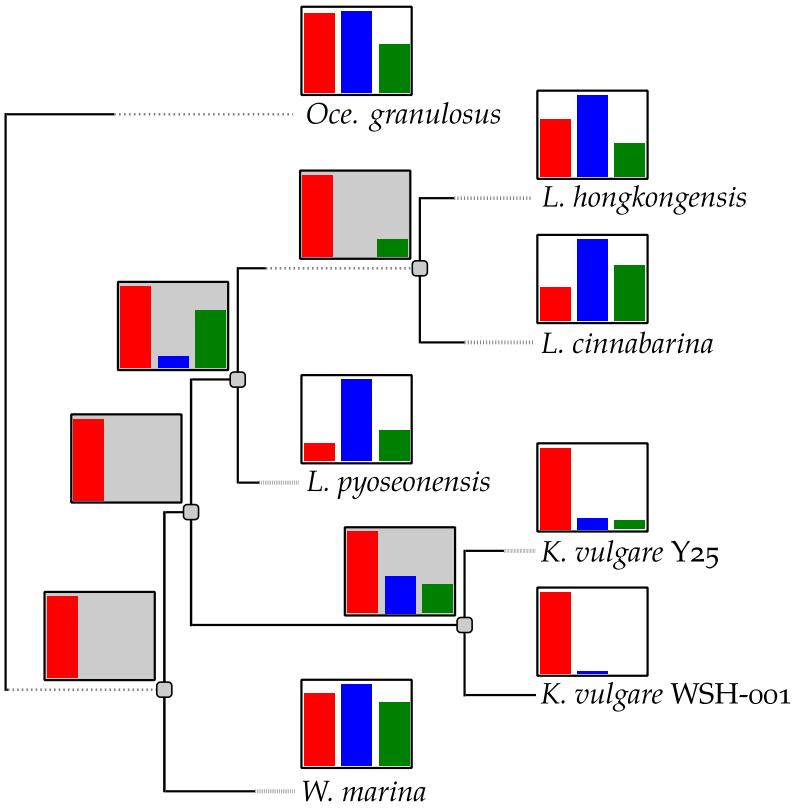
(a) overview

Figure A.7: Lineage-specific components of transfer (κ), loss (μ), duplication (λ), and edge length (t) parameters, as estimated by COUNT (Section 2.9). Trees present t directly (dotted lines do not count toward its value; scale is the same in all subfigures), while the other parameters are provided in bar charts. All parameters (including t) are normalised to μ, which therefore equals 1 in all charts. Charts at internal nodes are filled with grey background. Strain designations are omitted when appropriate, but are identical to those given in Table 2.2. (a) general overview, as well as parameters at deep nodes; (b) parameter values in the Loktanella clade; (c) parameter values in the smaller clade containing *Ketogulonicigenium*, some *Loktanella* species, and others.



(b) *Loktanella* genus

Figure A.7: Lineage-specific COUNT parameters (continued)



(c) *Ketogulonicigenium* and others

Figure A.7: Lineage-specific COUNT parameters (continued)

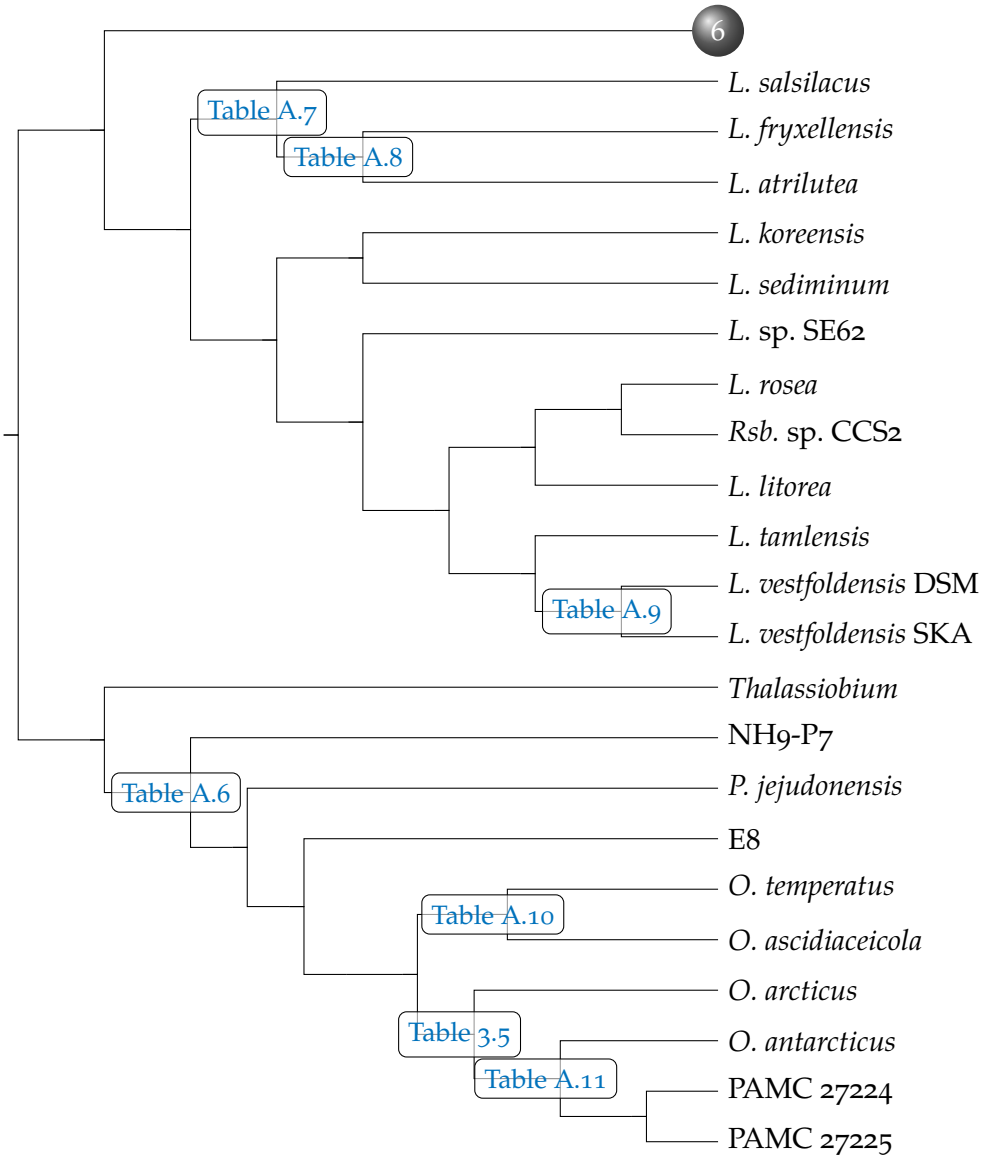


Figure A.8: Tables listing COG changes at specific nodes in Octadecabacter and Loktanella clade phylogeny. The information on gain and loss of specific COGs was used for comparison of changes associated with polar habitats (Section 4.3.4). *electronic version* – labels at internal nodes of the tree link directly to the respective table

Table A.6: Changes in COGs at the Octadecabacter clade members' LCA. Only COGs with a probability of change $p > 0.5$ at this node are listed. The approximate probability ratios are given in the rightmost column, with green and red indicating gain and loss, respectively. Relevant probabilities are gain (G) : expansion (E) : neutral (N, no change), and loss (L) : reduction (R) : neutral (N). A box indicates the highest probability, and is filled if it surpassed the second-highest by at least 50 %. The rightmost column gives the orders of magnitude between the highest and lowest ratio (the latter is always 1). COGs are sorted by category, with the respective category code given on the left. COGs assigned to more than one category are listed multiple times. (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
C	1454	Alcohol dehydrogenase, class IV	10	9	
G	3507	β-xylosidase	440	151	
H	1763	Molybdopterin-guanine dinucleotide biosynthesis protein	5033	783	
P	2998	ABC-type tungstate transport system, permease component	7705	882	
	4662	ABC-type tungstate transport system, periplasmic component	8562	1287	
R	3565	Predicted dioxygenase of extradiol dioxygenase family	141	4	10^3
	4277	Predicted DNA-binding protein with the Helix-hairpin-helix motif	66	4	10^3
S	4103	Uncharacterized conserved protein, tellurite resistance protein B (TerB) family	271	209	10^4
	4246	Uncharacterized protein	10	3	10^5
	4338	Uncharacterized protein	6672	1820	
U	0341	Preprotein translocase subunit SecF	29	1	10^4

Table A.7: Changes in COGs at the LCA of the two polar Loktanella isolates *L. fryxellensis* DSM 16213 and *L. salsilacus* DSM 16199. See description of Table A.6 for detailed information. (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
	0711	FoF1-type ATP synthase, membrane subunit b or b'	1	11	2	10^3

C

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0843	Heme/copper-type cytochrome/quinol oxidase, subunit 1	1	12	3	10^5
	1145	Ferredoxin	91	1	7	10^4
	1182	FMN-dependent NADH-azoreductase	2344	1	619	
	1301	Na^+/H^+-dicarboxylate symporter	23	1	2	10^5
	1454	Alcohol dehydrogenase, class IV	3	1	3	
	1622	Heme/copper-type cytochrome/quinol oxidase, subunit 2	1	168	58	10^4
	2326	Polyphosphate kinase 2, PPK2 family	1	3181	1878	
	2421	Acetamidase/formamidase	8118	1	3960	
	2838	Monomeric isocitrate dehydrogenase	24	1	2	10^3
	2864	Cytochrome b subunit of formate dehydrogenase	118	1	30	
	3278	Cbb3-type cytochrome oxidase, subunit 1	1	2328	1992	
D	1475	Chromosome segregation protein SpooJ, contains ParB-like nuclease domain	1	40	4	10^4
	3773	Cell wall hydrolase CwlJ, involved in spore germination	1	16	2	10^5
	0019	Diaminopimelate decarboxylase	1	24	9	10^4
	0069	Glutamate synthase domain 2	1	1886	582	
	0405	Gamma-glutamyltranspeptidase	1	29	14	10^3
	0686	Alanine dehydrogenase	1	428	48	
	0754	Glutathionylspermidine synthase	20	1	2	10^4
E	1003	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	5454	1	1034	
	1124	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	19	1	11	10^3

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	2066	Glutaminase	111	1	14	10^3
	3340	Peptidase E	147	1	8	10^3
	3591	V8-like Glu-specific endopeptidase	12	3	1	
	4302	Ethanolamine ammonia-lyase, small subunit	186	1	9	10^3
	4303	Ethanolamine ammonia-lyase, large subunit	105	1	3	10^4
	4311	Sarcosine oxidase delta subunit	1	576	25	
	4392	Branched-chain amino acid transport protein	456	1	140	
	4583	Sarcosine oxidase gamma subunit	1	299	14	
	0005	Purine nucleoside phosphorylase	1	1732	153	
F	0015	Adenylosuccinate lyase	1	16	11	10^4
	0788	Formyltetrahydrofolate hydrolase	11	1	9	10^3
	2233	Xanthine/uracil permease	12	1	1	10^3
	0296	1,4-alpha-glucan branching enzyme	3	2	1	
	0698	Ribose 5-phosphate isomerase RpiB	15	1	2	10^3
	1523	Pullulanase/glycogen debranching enzyme	2	1	1	
	1869	D-ribose pyranose/furanose isomerase RbsD	105	1	3	10^4
G	1877	Trehalose-6-phosphatase	6890	1	1035	
	1904	Glucuronate isomerase	519	1	498	
	2120	N-acetylglucosaminyl deacetylase, LmbE family	8700	1	4461	
	2133	Glucose/arabinose dehydrogenase, beta-propeller fold	1	123	27	
	2271	Sugar phosphate permease	38	1	8	10^3
	2513	2-Methylisocitrate lyase and related enzymes, PEP mutase family	1652	1	475	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	2861	Uncharacterized conserved protein YibQ, putative polysaccharide deacetylase 2 family	28	1	5	10^3
	2943	Membrane glycosyltransferase	493	1	433	
	3280	Maltooligosyltrehalose synthase	3345	1	829	
	3717	5-keto 4-deoxyuronate isomerase	530	1	177	
	4993	Glucose dehydrogenase	15	1	4	10^3
H	0161	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase	1	4134	787	
	1335	Nicotinamidase-related amidase	1	90	32	10^4
	2141	Flavin-dependent oxidoreductase, luciferase family (includes alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase)	1	56	51	10^4
	5598	Trimethylamine:corrinoid methyltransferase	1	117	13	
I	0236	Acyl carrier protein	1	26	19	10^3
	0657	Acetyl esterase/lipase	1	62	14	10^4
	0688	Phosphatidylserine decarboxylase	10	1	2	10^3
	0818	Diacylglycerol kinase	6504	1	777	
	1183	Phosphatidylserine synthase	1	163	13	10^4
	1597	Diacylglycerol kinase family enzyme	1200	1	626	
	2267	Lysophospholipase, alpha-beta hydrolase superfamily	1	177	39	10^3
	2854	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, periplasmic MlaC component	23	1	2	10^3
J	0590	tRNA(Arg) A34 adenosine deaminase TadA	1	309	32	10^4

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	1234	Ribonuclease BN, tRNA processing enzyme	9095	1	7188	
K	0782	Transcription elongation factor, GreA/GreB family	1	402	55	10^4
	1733	DNA-binding transcriptional regulator, HxlR family	91	1	35	
	1842	Phage shock protein A	30	1	4	10^3
	2183	Transcriptional accessory protein Tex/SPT6	7544	1	2231	
	2378	Predicted DNA-binding transcriptional regulator YafY, contains an HTH and WYL domains	1	2536	2019	
	3070	Transcriptional regulator of competence genes, TfoX/Sxy family	11	1	2	10^3
	4567	Two-component response regulator, ActR/RegA family, consists of REC and Fis-type HTH domains	1	340	25	10^4
	4941	Predicted RNA polymerase sigma factor, contains C-terminal TPR domain	3356	1	1221	
L	0420	DNA repair exonuclease SbcCD nuclease subunit	59	1	12	10^4
	0847	DNA polymerase III, epsilon subunit or related 3'-5' exonuclease	1	476	10	10^3
	1484	DNA replication protein DnaC	5146	1	25	
	1961	Site-specific DNA recombinase related to the DNA invertase Pin	1	3681	3178	
	4912	3-methyladenine DNA glycosylase AlkD	119	1	4	10^3
M	0381	UDP-N-acetylglucosamine 2-epimerase	618	1	176	
	0562	UDP-galactopyranose mutase	2455	1	2170	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	0767	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, permease component MlaE	1	75	31	
	1127	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, ATPase component MlaF	1	75	31	
	1210	UTP-glucose-1-phosphate uridylyltransferase	1	36	33	10^4
	1463	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, periplasmic component MlaD	1	75	31	
	1538	Outer membrane protein TolC	1	143	53	10^4
	2825	Periplasmic chaperone for outer membrane proteins, Skp family	1099	1	562	
	2853	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, lipoprotein component MlaA	22	1	3	10^3
	2943	Membrane glycosyltransferase	493	1	433	
	3307	O-antigen ligase	52	1	32	10^3
	3757	Lyzozyme M1 (1,4-beta-N-acetylmuramidase), GH25 family	28	1	4	10^3
	3773	Cell wall hydrolase CwlJ, involved in spore germination	1	164	17	10^4
N	1352	Methylase of chemotaxis methyl-accepting proteins	5	1	1	
	0501	Zn-dependent protease with chaperone function	13	1	13	
	0846	NAD-dependent protein deacetylase, SIR2 family	17	1	2	10^3
O	1404	Serine protease, subtilisin family	24	1	6	10^3
	1858	Cytochrome c peroxidase	1	794	94	
	2135	Putative SOS response-associated peptidase YedK	1	3234	321	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	2825	Periplasmic chaperone for outer membrane proteins, Skp family	1099	1	562	
	3187	Heat shock protein HslJ	658	1	288	
	3381	Cytoplasmic chaperone TorD involved in molybdoenzyme TorA maturation	677	1	348	
	4960	Flp pilus assembly protein, protease CpaA	995	1	201	
P	0053	Divalent metal cation (Fe/Co/Zn/Cd) transporter	25	1	19	
	0306	Phosphate/sulfate permease	144	1	8	10^2
	0569	Trk K^+ transport system, NAD-binding component	1	293	43	10^4
	0598	Mg^{2+} and Co^{2+} transporter CorA	27	1	25	
	0748	Putative heme iron utilization protein	16	1	3	10^3
	0753	Catalase	206	1	42	10^3
	1122	Energy-coupling factor transporter ATP-binding protein EcfA2	38	1	10	10^3
	1124	ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component	19	1	11	10^3
	1230	Co/Zn/Cd efflux system component	7	1	8	
	2608	Copper chaperone CopZ	2556	1	775	
	3119	Arylsulfatase A or related enzyme	1	1	1	
	3158	K^+ transporter	186	1	8	10^4
	3685	Ferritin-like metal-binding protein YciE	814	1	366	
Q	0236	Acyl carrier protein	1	26	19	10^3
	3485	Protocatechuate 3,4-dioxygenase beta subunit	7	1	2	
	3509	Poly(3-hydroxybutyrate) depolymerase	14	1	9	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	5285	Ectoine hydroxylase-related dioxygenase, phytanoyl-CoA dioxygenase (PhyH) family	353	1	62	
R	0121	Predicted glutamine amidotransferase	50	1	9	10^3
	0388	Predicted amidohydrolase	1	345	72	10^3
	0599	Uncharacterized conserved protein YurZ, alkylhydroperoxidase/carboxymuconolactone decarboxylase family	1	194	54	10^3
	1122	Energy-coupling factor transporter ATP-binding protein EcfA2	38	1	10	10^3
	1335	Nicotinamidase-related amidase	1	89	32	10^4
	1597	Diacylglycerol kinase family enzyme	1200	1	626	
	1611	Predicted Rossmann fold nucleotide-binding protein	1	9724	729	
	1741	Redox-sensitive bicupin YhaK, pirin superfamily	1	12	6	
	1917	Cupin domain protein related to quercetin dioxygenase	1	5060	1964	
	2124	Flavin-dependent oxidoreductase, luciferase family (includes alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase)	1	56	51	10^4
	2261	Uncharacterized membrane protein YeaQ/YmgE, transglycosylase-associated protein family	10	1	4	10^3
	2350	Uncharacterized conserved protein YciI, contains a putative active-site phosphohistidine	4879	1	130	
	3128	Predicted 2-oxoglutarate- and Fe(II)-dependent dioxygenase YbiX	78	1	26	10^3
	3360	Flavin-binding protein dodecin	8640	1	1541	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	3800	Predicted transcriptional regulator	7	22	1	
	4132	ABC-type uncharacterized transport system, permease component	105	1	3	10^4
	4589	Predicted CDP-diglyceride synthetase/phosphatidate cytidylyltransferase	32	1	2	10^3
	4666	TRAP-type uncharacterized transport system, fused permease components	1	21	20	
	4783	Putative Zn-dependent protease, contains TPR repeats	1	50	41	
	5496	Predicted thioesterase	105	1	3	10^4
	0586	Uncharacterized membrane protein DedA, SNARE-associated domain	5	9	1	
	2121	Uncharacterized conserved protein, lysophospholipid acyltransferase (LPLAT) superfamily	44	1	4	10^3
	2308	Uncharacterized conserved protein, circularly permuted ATPgrasp superfamily	1	4	1	
	2860	Uncharacterized membrane protein YeiH	23	1	5	10^3
	2898	Lysylphosphatidylglycerol synthetase, C-terminal domain, DUF2156 family	1	20	11	
	2979	Uncharacterized membrane protein YebE, DUF533 family	11	1	3	
S	3329	Uncharacterized conserved protein	1	1	2	
	3490	Uncharacterized protein	476	1	313	
	3514	Uncharacterized conserved protein, DUF4415 family	1099	1	302	
	3544	Uncharacterized conserved protein, DUF305 family	20	2	1	
	3603	Uncharacterized protein	66	1	60	10^3

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	3673	Uncharacterized protein, PA2063/DUF2235 family	1	48	33	
	3739	Uncharacterized membrane protein YoaT, DUF817 family	33	1	5	10^3
	3795	Uncharacterized conserved protein	5025	1	2271	
	3921	Uncharacterized conserved protein	1	1547	743	
	4196	Uncharacterized conserved protein, DUF2126 family	147	1	25	10^3
	4275	Uncharacterized protein	50	1	9	10^3
	4286	Uncharacterized protein, UPF0160 family	930	1	552	
	4298	Uncharacterized protein	30	1	4	10^3
	4307	Uncharacterized protein	228	1	17	10^3
	4325	Uncharacterized membrane protein	1	5	5	
	4405	Uncharacterized protein YhfF	1	297	135	
	4625	Uncharacterized conserved protein, contains a C-terminal beta-barrel porin domain	21	1	5	10^3
	4717	Uncharacterized protein YhaN	389	1	49	10^3
	4991	Uncharacterized conserved protein YraI	136	1	37	
	5463	Uncharacterized conserved protein YgiB, involved in bioifilm formation, UPF0441/DUF1190 family	111	1	8	10^3
	5481	Uncharacterized protein	10	1	1	10^3
	5579	Uncharacterized protein, DUF1810 family	1200	1	626	
	1352	Methylase of chemotaxis methyl-accepting proteins	5	1	1	
T	1842	Phage shock protein A	30	1	4	10^3
	3806	Anti-sigma factor ChrR, cupin superfamily	1	38	2	

Table A.7: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	4567	Two-component response regulator, ActR/RegA family, consists of REC and Fis-type HTH domains	1	340	25	10^4
	4960	Flp pilus assembly protein, protease CpaA	995	1	201	
V	1764	Organic hydroperoxide reductase OsmC/OhrA	3728	1	1584	
	4845	Chloramphenicol O-acetyltransferase	23	1	4	10^3
X	3547	Transposase	732	1	241	
	4584	Transposase	5756	1	3472	
	5534	Plasmid replication initiator protein	1432	1	448	

Table A.8: Changes in COGs at the LCA of the polar Loktanella isolate *L. fryxellensis* DSM 16213 and the temperate isolate *L. atrilutea* DSM 29326. See description of Table A.6 for detailed information. (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
G	3405	Endo-1,4-beta-D-glucanase Y	2768	1	2471	
J	4123	tRNA1(Val) A37 N6-methylase TrmN6	65	1	21	10^4
P	1910	Periplasmic molybdate-binding protein/domain	2820	1	1939	
	2216	High-affinity K^+ transport system, ATPase chain B	233	1	82	
R	2144	Selenophosphate synthetase-related protein	2654	1	1845	
	2516	Biotin synthase-related protein, radical SAM superfamily	2654	1	1845	
	3864	Predicted metal-dependent peptidase	3111	1	2120	
	4671	Predicted glycosyl transferase	216	1	85	
S	3506	Regulation of enolase protein 1 (function unknown), concanavalin A-like superfamily	4135	1	1944	

Table A.8: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N		
	3802	Uncharacterized protein	3183	1	2140
T	0467	RecA-superfamily ATPase, KaiC/GvpD/RAD55 family	2862	1	2540
U	4618	ABC-type protease/lipase transport system, ATPase and permease components	1	2016	1290

Table A.9: Changes in COGs at the LCA of *Loktanella vestfoldensis* DSM 16212 and SKA 53 (a polar and temperate isolate, respectively). See description of Table A.6 for detailed information. (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N		
C	0778	Nitroreductase	1	1844	150
	1182	FMN-dependent NADH-azoreductase	3781	1	708
	1251	NAD(P)H-nitrite reductase, large subunit	15	1	1 10^3
	1319	CO or xanthine dehydrogenase, FAD-binding subunit	1	135	107
	2326	Polyphosphate kinase 2, PPK2 family	1	3556	397
	3658	Cytochrome b	72	1	19 10^3
	4147	Na^+(or H^+)/acetate symporter ActP	16	1	9 10^3
D	1196	Chromosome segregation ATPase	1	8422	6645
	4942	Septal ring factor EnvC, activator of murein hydrolases AmiA and AmiB	1	4222	596
E	0069	Glutamate synthase domain 2	1	1345	304
	0308	Aminopeptidase N	455	1	6 10^4
	0509	Glycine cleavage system H protein (lipoate-binding)	61	1	2 10^3
	1003	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	45	1	5 10^3

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase	1	848	810	
	1247	L-amino acid N-acyltransferase YncA	45	1	2	10^3
	1410	Methionine synthase I, cobalamin-binding domain	1	176	21	10^4
	1770	Protease II	14	1	11	10^3
	1834	N-Dimethylarginine dimethylaminohydrolase	11	1	6	10^3
	2355	Zn-dependent dipeptidase, microsomal dipeptidase homolog	367	1	20	
	2755	Lysophospholipase L1 or related esterase	1	793	102	
	4160	ABC-type arginine/histidine transport system, permease component	1	268	102	
	4215	ABC-type arginine transport system, permease component	1	33	13	10^4
F	0563	Adenylate kinase or related kinase	1	6621	4071	
	0788	Formyltetrahydrofolate hydrolase	267	1	161	
G	0235	Ribulose-5-phosphate 4-epimerase/Fuculose-1-phosphate aldolase	48	1	3	
	0246	Mannitol-1-phosphate/altronate dehydrogenases	1	5269	328	
	0297	Glycogen synthase	29	1	22	
	0362	6-phosphogluconate dehydrogenase	188	1	3	10^4
	0366	Glycosidase	1	148	63	
	0448	ADP-glucose pyrophosphorylase	4533	1	3258	
	0580	Glycerol uptake facilitator and related aquaporins (Major Intrinsic Protein Family)	1032	1	805	

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0637	β-phosphoglucomutase or related phosphatase, HAD superfamily	221	1	15	
	0698	Ribose 5-phosphate isomerase RpiB	116	1	60	10^3
	2211	Na^+/melibiose symporter or related transporter	1	1278	724	
	2220	L-ascorbate metabolism protein UlaG, β-lactamase superfamily	38	1	3	
	2943	Membrane glycosyltransferase	80	1	53	
	3622	Hydroxypyruvate isomerase	145	1	6	
	4813	Trehalose utilization protein	55	1	43	10^3
H	0414	Panthothenate synthetase	72	1	2	10^3
	0432	Thiamin phosphate synthase YjbQ, UPF0047 family	37	1	29	10^3
	4032	Sulfopyruvate decarboxylase, TPP-binding subunit (coenzyme M biosynthesis)	69	1	37	10^3
I	0584	Glycerophosphoryl diester phosphodiesterase	1	6720	6281	
	3000	Sterol desaturase/sphingolipid hydroxylase, fatty acid hydroxylase superfamily	265	1	129	
	4850	Phosphatidate phosphatase APP1	142	1	88	10^3
J	3719	Ribonuclease I	369	1	5	10^4
K	0553	Superfamily II DNA or RNA helicase, SNF2 family	49	1	27	10^3
	0864	Metal-responsive transcriptional regulator, contains CopG/Arc/MetJ DNA-binding domain	55	1	9	10^4
	2186	DNA-binding transcriptional regulator, FadR family	1	8408	7368	
	3070	Transcriptional regulator of competence genes, TfoX/Sxy family	96	1	50	10^3

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
L	0350	O6-methylguanine-DNA–protein-cysteine methyltransferase	1	45	1	10^3
	0553	Superfamily II DNA or RNA helicase, SNF2 family	49	1	27	10^3
	1484	DNA replication protein DnaC	637	1	118	
	2176	DNA polymerase III, alpha subunit (gram-positive type)	7499	1	4127	
	2827	Predicted endonuclease, GIY-YIG superfamily	103	1	37	10^3
	3569	DNA topoisomerase IB	70	1	38	10^3
	4912	3-methyladenine DNA glycosylase AlkD	538	1	12	10^3
M	0399	dTDP-4-amino-4,6-dideoxygalactose transaminase	57	3	1	
	2943	Membrane glycosyltransferase	80	1	53	
	2989	Murein L,D-transpeptidase YcbB/YkuD	130	1	1	10^5
	5622	Protein required for attachment to host cells	105	1	58	10^3
O	0225	Peptide methionine sulfoxide reductase MsrA	1	102	2	10^3
	0846	NAD-dependent protein deacetylase, SIR2 family	100	1	51	10^3
	1858	Cytochrome c peroxidase	1	8159	4065	
	2020	Protein-S-isoprenylcysteine O-methyltransferase Ste14	1	136	110	
	2135	Putative SOS response-associated peptidase YedK	3258	1	97	
	2170	Gamma-glutamyl:cysteine ligase YbdK, ATP-grasp superfamily	14	1	10	10^3
	2370	Hydrogenase/urease accessory protein HupE	14	1	9	10^3
P	0310	ABC-type Co^{2+} transport system, permease component	298	1	34	10^3
	0475	Kef-type K^+ transport system, membrane component KefB	1	14	6	

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
	0704	Phosphate uptake regulator	1	37	13 10^4
	1814	Predicted Fe^{2+}/Mn^{2+} transporter, VIT1/CCC1 family	8121	1	7623
	1840	ABC-type Fe^{3+} transport system, periplasmic component	1	1569	108
	2116	Formate/nitrite transporter FocA, FNT family	88	1	48 10^3
	2215	ABC-type nickel/cobalt efflux system, permease component RcnA	33	1	31
	2608	Copper chaperone CopZ	2393	1	1421
	2847	Copper(I)-binding protein	1	12	9 10^3
	3197	Uncharacterized protein, possibly involved in nitrogen fixation	7344	1	480
	3221	ABC-type phosphate/phosphonate transport system, periplasmic component	1	8027	5756
	3454	α-D-ribose 1-methylphosphonate 5-triphosphate diphosphatase PhnM	4	1	1
	3720	Putative heme degradation protein	6749	1	5850
	4521	ABC-type taurine transport system, periplasmic component	69	1	52 10^3
	4558	ABC-type hemin transport system, periplasmic component	8006	1	7784
	4559	ABC-type hemin transport system, ATPase component	15	1	10 10^3
	4651	Predicted Kef-type K^+ transport protein, K^+/H^+ antiporter domain	199	1	36 10^3
Q	1647	Esterase/lipase	46	1	35 10^3
	2931	Ca^{2+}-binding protein, RTX toxin-related	33	1	1
	3509	Poly(3-hydroxybutyrate) depolymerase	40	1	40

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0384	Predicted epimerase YddE/YHI9, PhzF superfamily	1528	1	76	
	0637	β-phosphoglucomutase or related phosphatase, HAD superfamily	221	1	15	
	1063	Threonine dehydrogenase or related Zn-dependent dehydrogenase	1	848	810	
	1739	Putative translation regulator, IMPACT (imprinted ancient) protein family	31	1	3	10^3
R	1741	Redox-sensitive bicupin YhaK, pirin superfamily	1	3	3	
	1765	Uncharacterized OsmC-related protein	6585	1	5860	
	1917	Cupin domain protein related to quercetin dioxygenase	490	1	26	
	2249	Putative NADPH-quinone reductase (modulator of drug activity B)	26	1	18	
	2304	Secreted protein containing bacterial Ig-like domain and vWFA domain	48	1	4	10^3
	2321	Predicted metalloprotease	12	1	9	10^3
	3021	Uncharacterized conserved protein YafD, endonuclease/exonuclease/phosphatase (EEP) superfamily	336	1	216	
	3153	Predicted N-acetyltransferase YhbS	1	3007	118	
	3214	Uncharacterized conserved protein YcaQ, contains winged helix DNA-binding domain	7076	1	235	
	3217	Uncharacterized conserved protein YcbX, contains MOSC and Fe-S domains	67	1	40	10^3
	3360	Flavin-binding protein dodecin	116	1	66	10^3
	3380	Predicted NAD/FAD-dependent oxidoreductase	500	1	178	
	3393	Predicted acetyltransferase, GNAT family	47	1	5	10^3

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	3450	Predicted enzyme of the cupin superfamily	74	1	37	10^3
	3576	Predicted flavin-nucleotide-binding protein, pyridoxine 5′-phosphate oxidase superfamily	307	1	12	
	3828	Type 1 glutamine amidotransferase (GATase1)-like domain	65	1	11	10^3
S	1295	Uncharacterized membrane protein, BrkB/YihY/UPF0761 family (not an RNase)	1	895	480	
	2268	Uncharacterized membrane protein YqiK, contains Band7/PHB/SPFH domain	79	1	29	10^3
	2311	Uncharacterized membrane protein YeiB	62	1	17	10^3
	2833	Uncharacterized conserved protein, contains ferritin-like DUF455 domain	21	1	1	10^3
	3108	Uncharacterized conserved protein YcbK, DUF882 family	1	33	13	10^3
	3514	Uncharacterized conserved protein, DUF4415 family	21	1	15	10^3
	3544	Uncharacterized conserved protein, DUF305 family	249	1	187	
	3673	Uncharacterized protein, PA2063/DUF2235 family	1	434	236	
	3739	Uncharacterized membrane protein YoaT, DUF817 family	90	1	9	10^3
	3788	Uncharacterized membrane protein YecN, MAPEG domain	2783	1	1563	
	3832	Uncharacterized conserved protein YndB, AHSA1/START domain	1	5	2	
	3921	Uncharacterized conserved protein	1	28	2	
	4093	Uncharacterized protein	29	1	2	10^4
	4275	Uncharacterized protein	90	1	8	10^3

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	4327	Uncharacterized membrane protein	11	1	6	10^3
	4338	Uncharacterized protein	62	1	36	10^3
	4446	Uncharacterized conserved protein, DUF1499 family	78	1	48	10^3
	4944	Uncharacterized protein	111	1	28	10^3
	5375	Uncharacterized protein	38	1	28	10^3
	5425	Usg protein (tryptophan operon, function unknown)	103	1	56	10^3
	5453	Uncharacterized protein	76	1	4	10^3
	5470	Uncharacterized conserved protein, DUF1330 family	1	785	156	
T	2199	GGDEF domain, diguanylate cyclase (c-di-GMP synthetase) or its enzymatically inactive variants	1	128	60	
	2808	Predicted FMN-binding regulatory protein PaiB	74	1	35	10^3
	2905	Signal-transduction protein containing cAMP-binding, CBS, and nucleotidyltransferase domains	219	1	119	
	3045	Periplasmic catabolite regulation protein CreA (function unknown)	119	1	69	10^3
	3437	Response regulator c-di-GMP phosphodiesterase, RpfG family, contains REC and HD-GYP domains	9409	1	6369	
	3806	Anti-sigma factor ChrR, cupin superfamily	1	6359	2864	
	3916	N-acyl-L-homoserine lactone synthetase	1	120	38	
	4191	Signal transduction histidine kinase regulating C4-dicarboxylate transport system	1	859	82	
U	0848	Biopolymer transport protein ExbD	1	41	38	10^3

Table A.9: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N		
	5010	Flp pilus assembly protein TadD, contains TPR repeats	21	1	1
V	1566	Multidrug resistance efflux pump	1	44	43
	1680	CubicO group peptidase, β-lactamase class C family	1	389	249
W	5010	Flp pilus assembly protein TadD, contains TPR repeats	21	1	1
X	3436	Transposase	356	1	73

Table A.10: Changes in COGs at the LCA of *O. temperatus* and *O. ascidiaceicola*. See description of Table A.6 for detailed information. (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
C	0644	Dehydrogenase (flavoprotein)	1	5737	1964	
	0785	Cytochrome c biogenesis protein CcdA	1	60	5	10^3
	1049	Aconitase B	80	1	11	10^3
	1053	Succinate dehydrogenase/ fumarate reductase, flavoprotein subunit	1	71	34	10^3
	2041	Periplasmic DMSO/TMAO reductase YedYZ, molybdopterin-dependent catalytic subunit	1	15	8	10^3
	2055	Malate/lactate/ureidoglycolate dehydrogenase, LDH2 family	1	2559	1445	
	3258	Cytochrome c	31	1	10	10^3
	4106	Trans-aconitate methyltransferase	1	13	7	
	4736	Cbb3-type cytochrome oxidase, subunit 3	15	1	2	10^4
D	1475	Chromosome segregation protein SpoOJ, contains ParB-like nuclease domain	1	35	12	10^3
	4942	Septal ring factor EnvC, activator of murein hydrolases AmiA and AmiB	1	13	3	10^3

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	0112	Glycine/serine hydroxymethyltransferase	1	12	3	10^4
	0531	Amino acid transporter	13	1	3	10^4
E	1246	N-acetylglutamate synthase or related acetyltransferase, GNAT family	1362	1	724	
	1247	L-amino acid N-acyltransferase YncA	1	4	2	10^4
	1363	Putative aminopeptidase FrvX	43	1	1	10^4
	1748	Saccharopine dehydrogenase, NADP-dependent	2670	1	1505	
	2049	Allophanate hydrolase subunit 1	4280	1	2656	
	3191	L-aminopeptidase/D-esterase	17	1	2	10^3
	0297	Glycogen synthase	11	1	2	10^4
	0448	ADP-glucose pyrophosphorylase	11	1	1	10^4
	0574	Phosphoenolpyruvate synthase/ pyruvate phosphate dikinase	1	20	7	10^3
	0580	Glycerol uptake facilitator and related aquaporins (Major Intrinsic Protein Family)	1	17	5	
G	0726	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	32	1	20	
	0837	Glucokinase	1	55	7	10^5
	1363	Putative aminopeptidase FrvX	43	1	1	10^4
	1940	Sugar kinase of the NBD/HSP70 family, may contain an N-terminal HTH domain	1	8	4	
	2115	Xylose isomerase	25	1	4	10^5
	2211	Na^+/melibiose symporter or related transporter	98	10	1	
	2942	Mannose or cellobiose epimerase, N-acyl-D-glucosamine 2-epimerase family	18	1	8	10^3
	4124	β-mannanase	37	1	5	10^3
	4130	Predicted sugar epimerase, xylose isomerase-like family	6635	1	2434	

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	4213	ABC-type xylose transport system, periplasmic component	14	1	1	10^3
	4214	ABC-type xylose transport system, permease component	14	1	1	10^3
	4573	Tagatose-1,6-bisphosphate aldolase non-catalytic subunit AgaZ/GatZ	81	1	16	10^3
H	0054	6,7-dimethyl-8-ribityllumazine synthase (Riboflavin synthase β chain)	1	68	21	10^4
	0189	Glutathione synthase/ RimK-type ligase, ATP-grasp superfamily	1	2907	1358	
	0661	Predicted unusual protein kinase regulating ubiquinone biosynthesis, AarF/ABC1/UbiB family	1	80	5	10^3
	1120	ABC-type cobalamin/Fe^{3+}-siderophores transport system, ATPase component	31	1	8	10^3
	1335	Nicotinamidase-related amidase	1	14	5	10^4
	2141	Flavin-dependent oxidoreductase, luciferase family (includes alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase)	1	13	10	10^4
I	1562	Phytoene/squalene synthetase	66	2	1	
J	0189	Glutathione synthase/RimK-type ligase, ATP-grasp superfamily	1	2907	1358	
	1236	RNA processing exonuclease, beta-lactamase fold, Cft2 family	177	1	7	10^4
	1600	Epoxyqueuosine reductase QueG (queuosine biosynthesis)	1	50	8	10^5
K	1733	DNA-binding transcriptional regulator, HxlR family	7501	1	2950	

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	1940	Sugar kinase of the NBD/HSP70 family, may contain an N-terminal HTH domain	1	8	4	
	2315	Predicted DNA-binding protein with 'double-wing' structural motif, MmcQ/YjbR family	55	1	22	
	3284	Transcriptional regulator of acetoin/glycerol metabolism	13	1	4	10^3
	3609	Transcriptional regulator, contains Arc/MetJ-type RHH (ribbon-helix-helix) DNA-binding domain	59	1	12	10^3
	3829	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding Fis domains	9477	1	3704	
	4977	Transcriptional regulator GlxA family, contains an amidase domain and an AraC-type DNA-binding HTH domain	1	38	28	10^3
	0415	Deoxyribodipyrimidine photolyase	45	1	1	
	0582	Integrase	528	1	55	
L	0847	DNA polymerase III, ϵ subunit or related 3'-5' exonuclease	1	48	32	10^3
	1201	Lhr-like helicase	60	1	2	10^5
	1793	ATP-dependent DNA ligase	96	1	16	10^3
	2818	3-methyladenine DNA glycosylase Tag	1	50	13	10^4
	4973	Site-specific recombinase XerC	11	1	4	10^3
	0726	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	32	1	20	
	0767	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, permease component MlaE	1	1760	377	
M	1087	UDP-glucose 4-epimerase	1	9003	4981	
	1088	dTDP-D-glucose 4,6-dehydratase	1	17	3	

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
	1091	dTDP-4-dehydrorhamnose reductase	1	14	3
	1127	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, ATPase component MlaF	1	1760	377
	1209	dTDP-glucose pyrophosphorylase	1	10	2
	1463	ABC-type transporter Mla maintaining outer membrane lipid asymmetry, periplasmic component MlaD	1	1760	377
	1538	Outer membrane protein TolC	1	44	5 10^4
	2222	Fructoselysine-6-P-deglycase FrlB and related proteins with duplicated sugar isomerase (SIS) domain	55	1	27
	3524	Capsule polysaccharide export protein KpsE/RkpR	1	31	22 10^3
O	0443	Molecular chaperone DnaK (HSP70)	1	68	24 10^3
	0785	Cytochrome c biogenesis protein CcdA	1	60	5 10^3
	1305	Transglutaminase-like enzyme, putative cysteine protease	41	1	18
	1858	Cytochrome c peroxidase	1	50	5 10^3
	2143	Thioredoxin-related protein	6995	1	5018
	2170	Gamma-glutamyl:cysteine ligase YbdK, ATP-grasp superfamily	28	1	5 10^3
	3484	Predicted proteasome-type protease	98	1	3 10^4
P	0025	NhaP-type Na^+/H^+ or K^+/H^+ antiporter	39	1	2
	0288	Carbonic anhydrase	115	1	1 10^6
	0614	ABC-type Fe^{3+}-hydroxamate transport system, periplasmic component	8970	1	2220

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N		
	1120	ABC-type cobalamin/Fe^{3+}-siderophores transport system, ATPase component	31	1	8	10^3
	1629	Outer membrane receptor proteins, mostly Fe transport	362	1	45	
	2346	Truncated hemoglobin YjbI	1961	1	1119	
	2375	NADPH-dependent ferric siderophore reductase, contains FAD-binding and SIP domains	54	1	42	10^3
	3454	α-D-ribose 1-methylphosphonate 5-triphosphate diphosphatase PhnM	674	1	529	
	3720	Putative heme degradation protein	2658	1	1570	
	4558	ABC-type hemin transport system, periplasmic component	2156	1	1078	
	4559	ABC-type hemin transport system, ATPase component	20	1	6	10^3
	4604	ABC-type enterochelin transport system, ATPase component	42	1	10	10^3
	4605	ABC-type enterochelin transport system, permease component	9317	1	3566	
	4606	ABC-type enterochelin transport system, permease component	1735	1	990	
	4607	ABC-type enterochelin transport system, periplasmic component	4818	1	2015	
	4638	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, large terminal subunit	1	13	12	10^3
Q	1233	Phytoene dehydrogenase-related protein	276	1	63	
	2130	NADPH-dependent curcumin reductase CurA	27	1	23	10^4
	3191	L-aminopeptidase/D-esterase	17	1	2	10^3
R	1335	Nicotinamidase-related amidase	1	14	5	10^4
	1407	Metallophosphoesterase superfamily enzyme	60	1	2	10^5

Table A.10: continued (back to overview)

COG ID	DESCRIPTION	G:E:N	L:R:N		
1741	Redox-sensitive bicupin YhaK, pirin superfamily	1	277	144	
2130	NADPH-dependent curcumin reductase CurA	27	1	23	10^4
2141	Flavin-dependent oxidoreductase, luciferase family (includes alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase)	1	13	10	10^4
2350	Uncharacterized conserved protein YciI, contains a putative active-site phosphohistidine	1	15	2	10^3
3046	Uncharacterized protein related to deoxyribodipyrimidine photolyase	55	1	2	10^4
3218	ABC-type uncharacterized transport system, auxiliary component	20	1	3	10^4
3393	Predicted acetyltransferase, GNAT family	19	1	2	10^4
3453	Predicted phosphohydrolase, protein tyrosine phosphatase (PTP) superfamily, DUF442 family	1	18	8	10^3
3608	Predicted deacylase	4248	1	3419	
3828	Type 1 glutamine amidotransferase (GATase1)-like domain	20	1	3	10^4
4341	Predicted HD phosphohydrolase	2016	1	266	
4638	Phenylpropionate dioxygenase or related ring-hydroxylating dioxygenase, large terminal subunit	1	13	12	10^3
4782	Esterase/lipase superfamily enzyme	42	1	11	10^3
4922	Predicted SnoaL-like aldol condensation-catalyzing enzyme	58	1	18	10^3

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
S	1238	Uncharacterized membrane protein YqaA, SNARE-associated domain	120	1	1	10^4
	1432	Uncharacterized conserved protein, LabA/DUF88 family	1	1048	725	
	2307	Uncharacterized conserved protein, Alpha-E superfamily	21	1	1	10^4
	2308	Uncharacterized conserved protein, circularly permuted ATPgrasp superfamily	6328	1	2810	
	2862	Uncharacterized membrane protein YqhA	20	1	2	10^5
	3329	Uncharacterized conserved protein	17	1	3	
	3650	Uncharacterized membrane protein	2286	1	2050	
	3779	Uncharacterized conserved protein YegJ, DUF2314 family	52	1	7	10^4
	3813	Uncharacterized protein	30	1	5	10^4
	3932	Uncharacterized conserved protein	6297	1	3021	
	4283	Uncharacterized protein	63	1	4	10^4
	4337	Uncharacterized protein	72	1	5	10^4
	4427	Uncharacterized protein	15	1	1	10^4
	4852	Uncharacterized membrane protein	7114	1	3467	
	5375	Uncharacterized protein	9	1	6	10^4
	5477	Predicted small integral membrane protein	29	1	14	10^4
	5501	Predicted secreted protein	21	1	10	10^3
T	0661	Predicted unusual protein kinase regulating ubiquinone biosynthesis, AarF/ABC1/UbiB family	1	80	5	10^3
	0784	CheY chemotaxis protein or a CheY-like REC (receiver) domain	2153	1	194	
	3228	Mlc titration factor MtfA, regulates *ptsG* expression	72	1	5	10^4

Table A.10: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
	3806	Anti-sigma factor ChrR, cupin superfamily	1	42	10^3
	3829	Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding Fis domains	9477	1	3704
	3916	N-acyl-L-homoserine lactone synthetase	1	576	157
U	0342	Preprotein translocase subunit SecD	1	34	10^5
	0811	Biopolymer transport protein ExbB/TolQ	1	16	10^3
	0848	Biopolymer transport protein ExbD	1	17	10^4
	1826	Sec-independent protein translocase protein TatA	2994	1	523
	4618	ABC-type protease/lipase transport system, ATPase and permease components	1	11	10^3
X	0582	Integrase	528	1	55
	3668	Plasmid stabilization system protein ParE	812	1	760

Table A.11: Changes in COGs at the LCA of the *O. antarcticus* type strain and the two PAMC isolates. See description of Table A.6 for detailed information. (back to overview)

	COG ID	DESCRIPTION	G:E:N	L:R:N	
C	0651	Formate hydrogenlyase subunit 3/Multisubunit Na^+/H^+ antiporter, MnhD subunit	1	35	10^3
	1979	Alcohol dehydrogenase YqhD, Fe-dependent ADH family	5238	1	1850
E	0165	Argininosuccinate lyase	1	54	10^3
F	0647	Ribonucleotide monophosphatase NagD, HAD superfamily	1	20	10^4

Table A.11: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
G	1850	Ribulose 1,5-bisphosphate carboxylase, large subunit, or a RuBisCO-like protein	1166	1	1009	
	3734	2-keto-3-deoxy-galactonokinase	1	194	46	10^4
	3958	Transketolase, C-terminal subunit	1168	1	1019	
	3959	Transketolase, N-terminal subunit	1168	1	1019	
H	2896	Molybdenum cofactor biosynthesis enzyme MoaA	1	67	32	10^4
I	2230	Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases	1	15	12	10^3
K	1321	Mn-dependent transcriptional regulator, DtxR family	5128	1	2232	
	3279	DNA-binding response regulator, LytR/AlgR family	1	5	5	
	5662	Transmembrane transcriptional regulator (anti-sigma factor RsiW)	5442	1	2334	
M	1346	Putative effector of murein hydrolase	10	1	2	
	3713	Outer membrane scaffolding protein for murein synthesis, MipA/OmpV family	2	1	1	
P	0651	Formate hydrogenlyase subunit 3/Multisubunit Na^+/H^+ antiporter, MnhD subunit	1	35	14	10^3
	1320	Multisubunit Na^+/H^+ antiporter, MnhG subunit	1	15	13	10^3
	1863	Multisubunit Na^+/H^+ antiporter, MnhE subunit	1	15	13	10^3
S	3002	Uncharacterized conserved protein YbcC, UPF0753/DUF2309 family	9	1	2	
	3152	Uncharacterized membrane protein YhaH, DUF805 family	6	1	4	

Table A.11: continued (back to overview)

	COG ID	DESCRIPTION	G:E:N / L:R:N			
	3205	Uncharacterized membrane protein	6915	1	1791	
	3535	Uncharacterized conserved protein, DUF917 family	24	1	10	
	4728	Uncharacterized protein	3891	1	1533	
	5467	Uncharacterized protein	1	51	32	10^3
	5587	Uncharacterized conserved protein, DUF2461 family	2	1	2	
T	3279	DNA-binding response regulator, LytR/AlgR family	1	5	5	
U	0811	Biopolymer transport protein ExbB/TolQ	1	11	5	10^3
	0848	Biopolymer transport protein ExbD	1	70	32	10^3
V	0610	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases	106	1	96	
X	3654	Prophage maintenance system killer protein	1104	1	971	

A.7 READ COVERAGE OF PAMC 27224 CONTIGS

To assess variances in genome coverage, the sequencing reads of isolate PAMC 27224 were mapped back onto the 132 assembled contigs using `bowtie2`. Coverage of each individual contig was related to the median coverage (Figure A.9). A high ratio can indicate the presence of multiple copies of genes or features within that particular contig. While the method deployed here is less sophisticated than other described workflows (e. g. [344]), its purpose is only to serve as a rough estimate. Among the contigs with the highest coverage (≈6-fold higher than the median coverage) is the one containing the rRNA genes. This indicates that these are present in multiple copies throughout the genome, although they are annotated just once, likely due to the high fragmentation. Other contigs with high coverage usually contain one or more transposase-encoding genes, which could mean that their number, too, is underestimated from the annotated CDSs.

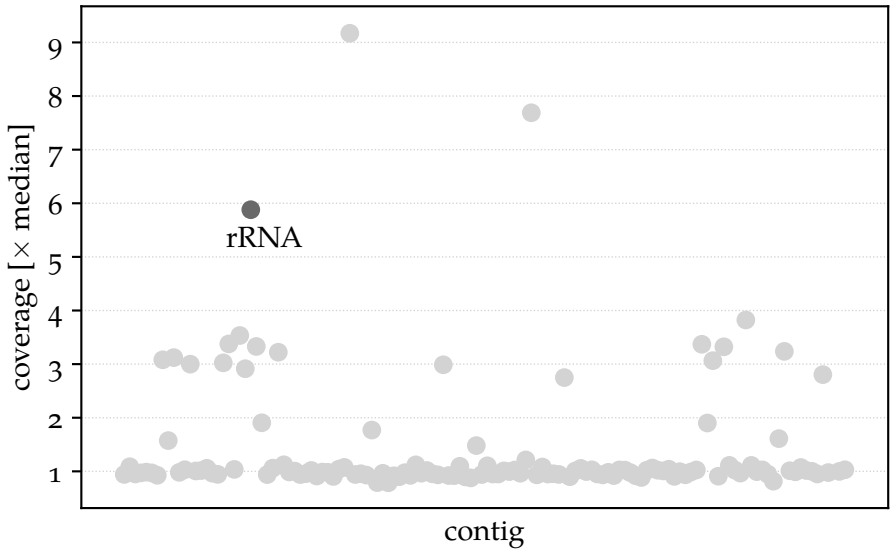


Figure A.9: Read coverage of isolate PAMC 27224 contigs. Reads were mapped back onto the genome using `bowtie2`, and the coverage for each contig was divided by the median coverage of all contigs. Contigs with a coverage of 3- to 9-fold the median exist, and usually encode one or multiple transposases. In addition, the contig carrying the rRNA genes shows a high coverage (highlighted in the plot), which indicates multiple copies of these genes in the genome.

BIBLIOGRAPHY

[1] Matthew A. Charette and Walter H.F. Smith. "The volume of earth's ocean." In: *Oceanography* 23.2 (June 2010), pp. 112–114. DOI: 10.5670/oceanog.2010.51.

[2] A. Baumgartner and E. Reichel. *The world water balance: mean annual global, continental and maritime precipitation, evaporation and run-off*. Elsevier Scientific Publishing Company, 1975. ISBN: 9780444998583.

[3] Robert F. Adler et al. "The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present)." In: *Journal of Hydrometeorology* 4.6 (2003), pp. 1147–1167. DOI: 10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

[4] Paul Durack. "Ocean salinity and the global water cycle." In: *Oceanography* 28.1 (2015), pp. 20–31. DOI: 10.5670/oceanog.2015.03.

[5] J. I. Hedges and J. M. Oades. "Comparative organic geochemistries of soils and marine sediments." In: *Organic Geochemistry* 27.7-8 (1997), pp. 319–361. DOI: 10.1016/S0146-6380(97)00056-9.

[6] Philippe Ciais et al. "Carbon and other biogeochemical cycles." In: *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley. Cambridge: Cambridge University Press, 2013. Chap. 6, pp. 465–570. ISBN: 978-1-107-66182-0. DOI: 10.1017/CBO9781107415324.015.

[7] Christopher B Field, Michael J Behrenfeld, James T Randerson, and Paul Falkowski. "Primary production of the biosphere: integrating terrestrial and oceanic components." In: *Science* 281.5374 (1998), pp. 237–240. DOI: 10.1126/science.281.5374.237.

[8] Mary Ann Moran et al. "Deciphering ocean carbon in a changing world." In: *Proceedings of the National Academy of Sciences* 113.12 (2016), pp. 3143–3151. DOI: 10.1073/pnas.1514645113.

[9] W B Whitman, D C Coleman, and W J Wiebe. "Prokaryotes: the unseen majority." In: *Proceedings of the National Academy of Sciences of the United States of America* 95.12 (1998), pp. 6578–83. DOI: 10.1073/pnas.95.12.6578.

[10] Marcel Margulies et al. "Genome sequencing in microfabricated high-density picolitre reactors." In: *Nature* 437.7057 (2005), pp. 376–380. DOI: 10.1038/nature03959.

[11] Gerardo Turcatti, Anthony Romieu, Milan Fedurco, and Ana Paula Tairi. "A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis." In: *Nucleic Acids Research* 36.4 (2008). DOI: 10.1093/nar/gkn021.

[12] Ângela M. Ribeiro, Andrew D. Foote, Anne Kupczok, Bárbara Frazão, Morten T. Limborg, Rosalía Piñeiro, Samuel Abalde, Sara Rocha, and Rute R. da Fonseca. "Marine genomics: News and views." In: *Marine Genomics* 31 (2017), pp. 1 –8. DOI: https://doi.org/10.1016/j.margen.2016.09.002.

[13] Sonja Voget et al. "Adaptation of an abundant Roseobacter RCA organism to pelagic systems revealed by genomic and transcriptomic analyses." In: *The ISME Journal* 9.2 (2015), pp. 371–384. DOI: 10.1038/ismej.2014.134.

[14] Soo Je Park, Rohit Ghai, Ana Belén Martín-Cuadrado, Francisco Rodríguez-Valera, Won Hyong Chung, Kae Kyoung Kwon, Jung Hyun Lee, Eugene L. Madsen, and Sung Keun Rhee. "Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments." In: *PLoS ONE* 9.5 (2014), pp. 1–10. DOI: 10.1371/journal.pone.0096449.

[15] B A Methé et al. "The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses." In: *Proceedings of the National Academy of Sciences of the United States of America* 102.31 (2005), pp. 10913–10918. DOI: Doi10.1073/Pnas.0504766102.

[16] Nicola Segata, Daniela Boernigen, Timothy L. Tickle, Xochitl C. Morgan, Wendy S. Garrett, and Curtis Huttenhower. "Computational meta'omics for microbial community studies." In: *Molecular Systems Biology* 9.1 (2013), pp. 1–15. DOI: 10.1038/msb.2013.22.

[17] R. Chester and T.D. Jickells. *Marine Geochemistry*. Wiley, 2012. ISBN: 9781118349090.

[18] C. Wuchter et al. "Archaeal nitrification in the ocean." In: *Proceedings of the National Academy of Sciences* 103.33 (2006), pp. 12317–12322. DOI: 10.1073/pnas.0600756103.

[19] Brandon K Swan et al. "Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean." In: *Science* 333.September (2011), pp. 1296–9. DOI: 10.1126/science.1203690.

[20] Mikhail V. Zubkov and Manuela Hartmann. "Ecological significance of microbial trophic mixing in the oligotrophic ocean: the Atlantic Ocean case studies." In: *Microbial Ecology of the Oceans*. Ed. by Josep M. Gasol and David L. Kirchman. 3. Hoboken, NJ: Wiley Blackwell, 2018. Chap. 3, pp. 99–122. ISBN: 9781119107194.

[21] Author R a Duce et al. "Impacts of atmospheric Nitrogen on the anthropogenic open ocean." In: *Science* 320.5878 (2008), pp. 893–897. DOI: 10.1126/science.1150369.

[22] Jed A. Fuhrman and Åke Hagström. "Bacterial and archaeal community structure and its patterns." In: *Microbial Ecology of the Oceans*. Ed. by David L. Kirchman. 2. Hoboken, NJ: Wiley Blackwell, 2008. Chap. 3, pp. 45–90. ISBN: 978-0-470-04344-8.

[23] Stephen J. Giovannoni et al. "Genome streamlining in a cosmopolitan oceanic bacterium." In: *Science* 309.5738 (2005), pp. 1242–1245. DOI: 10.1126/science.1114057.

[24] Alison Buchan, José M. González, and Mary Ann Moran. "Overview of the marine *Roseobacter* lineage." In: *Applied and Environmental Microbiology* 71.10 (2005), pp. 5665–5677. DOI: 10.1128/AEM.71.10.5665-5677.2005.

[25] Josep M. Gasol and David L. Kirchman. "Introduction: The evolution of microbial ecology of the ocean." In: *Microbial Ecology of the Oceans*. Ed. by Josep M. Gasol and David L. Kirchman. 3. Hoboken, NJ: Wiley Blackwell, 2018. Chap. 1, pp. 1–46. ISBN: 9781119107194.

[26] Nicholas J. Russell. "Psychrophilic bacteria - Molecular adaptations of membrane lipids." In: *Comparative Biochemistry and Physiology - A Physiology* 118.3 (1997), pp. 489–493. DOI: 10.1016/S0300-9629(97)87354-9.

[27] Salvino D'Amico, Tony Collins, Jean-Claude Marx, Georges Feller, and Charles Gerday. "Psychrophilic microorganisms: Challenges for life." In: *EMBO reports* 7.4 (2006), pp. 385–9. DOI: 10.1038/sj.embor.7400662.

[28] Charles Gerday. "Psychrophily and catalysis." In: *Biology (Basel)* 2.2 (2013), pp. 719–741. DOI: 10.3390/biology2020719.

[29] Khawar Sohail Siddiqui and Ricardo Cavicchioli. "Cold-adapted enzymes." In: *Annual Review of Biochemistry* 75 (2006), pp. 403–433. DOI: 10.1146/annurev.biochem.75.103004.142723.

[30] G. Horn, R. Hofweber, W. Kremer, and H. R. Kalbitzer. "Structure and function of bacterial cold shock proteins." In: *Cellular and Molecular Life Sciences* 64.12 (2007), pp. 1457–1470. DOI: 10.1007/s00018-007-6388-4.

[31] Christian Lorenz, Christina E. Lünse, and Mario Mörl. "tRNA modifications: Impact on structure and thermal adaptation." In: *Biomolecules* 7.2 (2017), p. 35. DOI: 10.3390/biom7020035.

[32] Naoki Sato, Takashi Tachikawa, Akira Wada, and Ayumi Tanaka. "Temperature-dependent regulation of the ribosomal small-subunit protein S21 in the cyanobacterium *Anabaena variabilis* M3." In: *Journal of Bacteriology* 179.22 (1997), pp. 7063–7071. DOI: 10.1128/jb.179.22.7063-7071.1997.

[33] Kevin P. O'Connell and Michael F. Thomashow. "Transcriptional organization and regulation of a polycistronic cold shock operon in *Sinorhizobium meliloti* RM1021 encoding homologs of the *Escherichia coli* major cold shock gene *cspA* and ribosomal protein gene *rpsU*." In: *Applied and Environmental Microbiology* 66.1 (2000), pp. 392–400. DOI: 10.1128/AEM.66.1.392-400.2000.

[34] Jason P. Chin et al. "Solutes determine the temperature windows for microbial survival and growth." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.17 (2010), pp. 7835–7840. DOI: 10.1073/pnas.1000557107.

[35] Corien Bakermans, Peter W. Bergholz, Debora F. Rodrigues, Tatiana A. Vishnivetskaya, Héctor L. Ayala-del Río, and James M. Tiedje. "Genomic and expression analyses of cold-adapted microorganisms." In: *Polar Microbiology: Life in a Deep Freeze*. Ed. by Robert V Miller and Lyle G Whyte. 1. ASM Press, 2012. Chap. 6, pp. 126–155. ISBN: 978-1-55581-604-9. DOI: 10.1128/9781555817183.

[36] S Venketesh and C Dayananda. "Properties, potentials, and prospects of antifreeze proteins." In: *Critical Reviews in Biotechnology* 28.1 (2008), pp. 57–82. DOI: 10.1080/07388550801891152.

[37] C. A. Mancuso Nichols, J. Guezennec, and J. P. Bowman. "Bacterial exopolysaccharides from extreme marine environments with special consideration of the Southern Ocean, sea ice, and deep-sea hydrothermal vents: A review." In: *Marine Biotechnology* 7.4 (2005), pp. 253–271. DOI: 10.1007/s10126-004-5118-2.

[38] C. Krembs, H. Eicken, K. Junge, and J. W. Deming. "High concentrations of exopolymeric substances in Arctic winter sea ice: Implications for the polar ocean carbon cycle and cryoprotection of *Diatoms*." In: *Deep-Sea Research Part I: Oceanographic Research Papers* 49.12 (2002), pp. 2163–2181. DOI: 10.1016/S0967-0637(02)00122-X.

[39] Claudine Médigue et al. "Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125." In: *Genome Research* 15.10 (2005), pp. 1325–1335. DOI: 10.1101/gr.4126905.

[40] Jacqueline Goordial et al. "Cold adaptive traits revealed by comparative genomic analysis of the eurypsychrophile *Rhodococcus* sp. JG3 isolated from high elevation McMurdo Dry Valley permafrost, Antarctica." In: *FEMS Microbiology Ecology* 92.2 (2016), pp. 1–11. DOI: 10.1093/femsec/fiv154.

[41] Kevin R Arrigo. "Sea ice ecosystems." In: *Annu. Rev. Marine. Sci* 6 (2014), pp. 439–467. DOI: 10.1146/annurev-marine-010213-135103.

[42] Antje Boetius, Alexandre M Anesio, Jody W Deming, Jill Mikucki, and Josephine Z Rapp. "Microbial ecology of the cryosphere: sea ice and glacial habitats." In: *Nature Reviews Microbiology* 13.11 (2014), pp. 677–690. DOI: 10.1038/nrmicro3522.

[43] Christopher Krembs, Hajo Eicken, and Jody W Deming. "Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic." In: *Proceedings of the National Academy of Sciences* 108.9 (2011), pp. 3653–3658. DOI: 10.1073/pnas.1100701108.

[44] Marcela Ewert and Jody Deming. "Sea ice microorganisms: Environmental constraints and extracellular responses." In: *Biology (Basel)* 2.2 (2013), pp. 603–628. DOI: 10.3390/biology2020603.

[45] David N. Thomas, Stathys Papadimitriou, and Christine Michel. "Biogeochemistry of sea ice." In: *Sea Ice*. Ed. by David N. Thomas and Gerhard S. Diekmann. 2. Wiley-Blackwell, 2010. Chap. 12, pp. 425–467. ISBN: 978-1-4051-8580-6.

[46] R. Bargagli. "Environmental contamination in Antarctic ecosystems." In: *Science of the Total Environment* 400.1-3 (2008), pp. 212–226. DOI: 10.1016/j.scitotenv.2008.06.062.

[47] K. Van de Velde, P. Vallelonga, J. P. Candelone, K. J.R. Rosman, V. Gaspari, G. Cozzi, C. Barbante, R. Udisti, P. Cescon, and C. F. Boutron. "Pb isotope record over one century in snow from Victoria Land, Antarctica." In: *Earth and Planetary Science Letters* 232.1-2 (2005), pp. 95–108. DOI: 10.1016/j.epsl.2005.01.007.

[48] Ralf Ebinghaus, Hans H. Kock, Christian Temme, Jürgen W. Einax, Astrid G. Löwe, Andreas Richter, John P. Burrows, and William H. Schroeder. "Antarctic springtime depletion of atmospheric mercury." In: *Environmental Science and Technology* 36.6 (2002), pp. 1238–1244. DOI: 10.1021/es015710z.

[49] Evert Nieboer and David H S Richardson. "The replacement of the nondescript term 'heavy metals' by a biologically and chemically significant classification of metal ions." In: *Environmental Pollution (Series B)* 1 (1980), pp. 3–26.

[50] S Madronich, R L McKenzie, L O Björn, and M M Caldwell. "Changes in biologically active ultraviolet radiation reaching the Earth's surface." In: *Journal of photochemistry and photobiology. B: Biology* 46.1-3 (1998), pp. 5–19. DOI: 10.1039/b700017k.

[51] Robin Brinkmeyer, Katrin Knittel, Jutta Jürgens, Horst Weyland, Rudolf Amann, and Elisabeth Helmke. "Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice." In: *Applied and Environmental Microbiology* 69.11 (2003), pp. 6610–6619. DOI: 10.1128/AEM.69.11.6610-6619.2003.

[52] Jeff S Bowman, Simon Rasmussen, Nikolaj Blom, Jody W Deming, Søren Rysgaard, and Thomas Sicheritz-Ponten. "Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene." In: *The ISME journal* 6.1 (2012), pp. 11–20. DOI: 10.1038/ismej.2011.76.

[53] Rebecca O M Cowie, Gareth J. Williams, Elizabeth W. Maas, K. Matt Voyles, and Ken G. Ryan. "Antarctic sea-ice microbial communities show distinct patterns of zonation in response to algal-derived substrates." In: *Aquatic Microbial Ecology* 73.2 (2014), pp. 123–134. DOI: 10.3354/ame01710.

[54] M. Hünken, J. Harder, and G. O. Kirst. "Epiphytic bacteria on the Antarctic ice diatom *Amphiprora kufferathii* Manguin cleave hydrogen peroxide produced during algal photosynthesis." In: *Plant Biology* 10.4 (2008), pp. 519–526. DOI: 10.1111/j.1438-8677.2008.00040.x.

[55] Stefanie VanTrappen, Joris Mergaert, and Jean Swings. "*Loktanella salsilacus* gen. nov., sp. nov., *Loktanella fryxellensis* sp. nov. and *Loktanella vestfoldensis* sp. nov., new members of the Rhodobacter group isolated from microbial mats in Antarctic lakes." In: *International Journal of Systematic and Evolutionary Microbiology* 54.4 (2004), pp. 1263–1269. DOI: 10.1099/ijs.0.03006-0.

[56] De Chao Zhang, Hui Rong Li, Yu Hua Xin, Hong Can Liu, Zhen Ming Chi, Pei Jin Zhou, and Yong Yu. "*Phaeobacter arcticus* sp. nov., a psychrophilic bacterium isolated from the Arctic." In: *International Journal of Systematic and Evolutionary Microbiology* 58.6 (2008), pp. 1384–1387. DOI: 10.1099/ijs.0.65708-0.

[57] Tsuneo Shiba. "*Roseobacter litoralis* gen. nov., sp. nov., and *Roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll a." In: *Systematic and Applied Microbiology* 14.2 (1991), pp. 140–145. DOI: 10.1016/S0723-2020(11)80292-4.

[58] Keizo Shimada. "Aerobic anoxygenic phototrophs." In: *Anoxygenic Photosynthetic Bacteria*. Ed. by Robert E. Blankenship, Michael T. Madigan, and Carl E. Bauer. Dordrecht: Kluwer Academic Publishers, 1995. Chap. 6, pp. 105–122. ISBN: 0-7923-3681-X.

[59] Meinhard Simon, Carmen Scheuner, Jan P. Meier-Kolthoff, Thorsten Brinkhoff, Irene Wagner-Döbler, Marcus Ulbrich, Hans-Peter Klenk, Dietmar Schomburg, Jörn Petersen, and Markus Göker. "Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats." In: *The ISME Journal* 11.6 (2017), pp. 1483–1499. ISSN: 1751-7362. DOI: 10.1038/ismej.2016.198.

[60] Ryan J Newton et al. "Genome characteristics of a generalist marine bacterial lineage." In: *The ISME journal* 4.6 (2010), pp. 784–798. DOI: 10.1038/ismej.2009.150.

[61] John Vollmers. "Molekularbiologische Charakterisierung und vergleichende Genomik von ausgewählten Vertretern mariner *Roseobacter*-Stämme." dissertation. Georg-August Universität Göttingen, 2013. URL: \url{https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-0022-5E61-6?locale-attribute=en}.

[62] Haiwei Luo and Mary Ann Moran. "Evolutionary ecology of the marine Roseobacter clade." In: *Microbiology and Molecular Biology Reviews* 78.4 (2014), pp. 573–587. DOI: 10.1128/MMBR.00020-14.

[63] Ying Sun, Kate E. Powell, Way Sung, Michael Lynch, Mary Ann Moran, and Haiwei Luo. "Spontaneous mutations of a model heterotrophic marine bacterium." In: *ISME Journal* 11.7 (2017), pp. 1713–1718. DOI: 10.1038/ismej.2017.20.

[64] Hui Wang, Jürgen Tomasch, Michael Jarek, and Irene Wagner-Döbler. "A dual-species co-cultivation system to study the interactions between Roseobacters and dinoflagellates." In: *Frontiers in Microbiology* 5.JUN (2014), p. 311. DOI: 10.3389/fmicb.2014.00311.

[65] Bryndan P. Durham et al. "Cryptic carbon and sulfur cycling between surface ocean plankton." In: *Proceedings of the National Academy of Sciences* 112.2 (2015), pp. 453–457. DOI: 10.1073/pnas.1413137112.

[66] Elke Allers, Laura Gómez-Consarnau, Jarone Pinhassi, Josep M. Gasol, Karel Šimek, and Jakob Pernthaler. "Response of *Alteromonadaceae* and *Rhodobacteriaceae* to glucose and phosphorus manipulation in marine mesocosms." In: *Environmental Microbiology* 9.10 (2007), pp. 2417–2429. DOI: 10.1111/j.1462-2920.2007.01360.x.

[67] Jose M Gonzalez, Ronald P Kiene, and Mary Ann Moran. "Transformation of sulfur compounds by an abundant lineage of marine bacteria in the α-subclass of the class Proteobacteria." In: *Applied and Environmental Microbiology* 65.9 (1999), pp. 3810–3819.

[68] Jingli Liu, Ji Liu, Sheng-Hui Zhang, Jinchang Liang, Heyu Lin, Delei Song, Gui-Peng Yang, Jonathan D. Todd, and Xiao-Hua Zhang. "Novel insights into bacterial dimethylsulfoniopropionate catabolism in the East China Sea." In: *Frontiers in Microbiology* 9.December (2018), p. 3206. DOI: 10.3389/fmicb.2018.03206.

[69] Martin F. Polz, Dana E. Hunt, Sarah P. Preheim, and Daniel M. Weinreich. "Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 361.1475 (2006), pp. 2009–2021. DOI: 10.1098/rstb.2006.1928.

[70] M. A. Moran et al. "Ecological genomics of marine roseobacters." In: *Applied and Environmental Microbiology* 73.14 (2007), pp. 4559–4569. DOI: 10.1128/AEM.02580-06.

[71] Marc Solioz and Barry Marrs. "The gene transfer agent of *Rhodopseudomonas capsulata*: Purification and characterization of its nucleic acid." In: *Archives of Biochemistry and Biophysics* 181 (1977), pp. 300–307. DOI: 10.1016/0003-9861(77)90508-2.

[72] Erin J. Biers, Kui Wang, Catherine Pennington, Robert Belas, Feng Chen, and Mary Ann Moran. "Occurrence and expression of gene transfer agent genes in marine bacterioplankton." In: *Applied and Environmental Microbiology* 74.10 (2008), pp. 2933–2939. DOI: 10.1128/AEM.02129-07.

[73] Catalina Maria Alejandro-Marín, Rafael Bosch, and Balbina Nogales. "Comparative genomics of the protocatechuate branch of the β-ketoadipate pathway in the *Roseobacter* lineage." In: *Marine Genomics* 17 (2014), pp. 25–33. DOI: 10.1016/j.margen.2014.05.008.

[74] Pascal Bartling, John Vollmers, and Jörn Petersen. "The first world swimming championships of roseobacters-—Phylogenomic insights into an exceptional motility phenotype." In: *Systematic and Applied Microbiology* 41.6 (2018), pp. 544–554. DOI: 10.1016/j.syapm.2018.08.012.

[75] Michal Koblížek, Qiang Zheng, Aleš Horák, and Miroslav Oborník. "Regressive evolution of photosynthesis in the Roseobacter clade." In: *Advances in Botanical Research Volume 66 – Genome Evolution of Photosynthetic Bacteria*. Ed. by Jean-pierre Jacquot, Pierre Gadal, and J. Thomas Beatty. Amsterdam: Elsevier Ltd., 2013. Chap. 13, pp. 386–405. ISBN: 9780123979230. DOI: 10.1016/B978-0-12-397923-0.00012-6.

[76] J.J. Gosink, R.P. Herwig, and J.T. Staley. "*Octadecabacter arcticus* gen. nov., sp. nov., and *O. antarcticus*, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water." In: *Systematic and Applied Microbiology* 20.3 (1997), pp. 356–365. DOI: 10.1016/S0723-2020(97)80003-3.

[77] John Vollmers, Sonja Voget, Sascha Dietrich, Kathleen Gollnow, Maike Smits, Katja Meyer, Thorsten Brinkhoff, Meinhard Simon, and Rolf Daniel. "Poles apart: Arctic and Antarctic *Octadecabacter* strains share high genome plasticity and a new type of xanthorhodopsin." In: *PloS one* 8.5 (2013), e63422. DOI: 10.1371/journal.pone.0063422.

[78] Sooyeon Park and Jung Hoon Yoon. "*Octadecabacter jejudonensis* sp. nov., isolated from the junction between the ocean and a freshwater spring and emended description of the genus *Octadecabacter*." In: *International Journal of Systematic and Evolutionary Microbiology* 64.PART 3 (2014), pp. 719–724. DOI: 10.1099/ijs.0.057513-0.

[79] Sara Billerbeck, Julia Orchard, Brian J. Tindall, Helge Ansgar Giebel, Thorsten Brinkhoff, and Meinhard Simon. "Description of *Octadecabacter temperatus* sp. nov., isolated from the southern North Sea, emended descriptions of the genus *Octadecabacter* and its species and reclassification of *Octadecabacter jejudonensis* Park and Yoon 2014 as *Pseudooctadecabacter jejudonensis* gen. nov comb. nov." In: *International Journal of Systematic and Evolutionary Microbiology* 65.6 (2015), pp. 1967–1974. DOI: 10.1099/ijs.0.000205.

[80] Yung Mi Lee, You-jung Jung, Soon Gyu Hong, Ji Hee Kim, and Hong Kum Lee. "Diversity and physiological characteristics of culturable bacteria from marine sediments of Ross Sea, Antarctica." In: *Korean Journal of Microbiology* 50.2 (2014), pp. 119–127.

[81] Young Ok Kim, In Suk Park, Sooyeon Park, Bo Hye Nam, Ji Min Park, Dong Gyun Kim, and Jung Hoon Yoon. "*Octadecabacter ascidiaceicola* sp. nov., isolated from a sea squirt (*Halocynthia roretzi*)." In: *International Journal of Systematic and Evolutionary Microbiology* 66.1 (2016), pp. 296–301. DOI: 10.1099/ijsem.0.000715.

[82] Sooyeon Park, Sun Young Yoon, Yong-taek Jung, and Jung-hoon Yoon. "*Octadecabacter ponticola* sp. nov., isolated from seawater." In: *International Journal of Systematic and Evolutionary Microbiology* 66.10 (2016), pp. 4179–4184. DOI: 10.1099/ijsem.0.001332.

[83] Sven Breider, Silke Pradella, Carmen Scheuner, Jörn Petersen, Peter Schumann, Hans-Peter Klenk, Thorsten Brinkhoff, Markus Göker, and Anne Fiebig. "Genome-scale data suggest reclassifications in the Leisingera-Phaeobacter cluster including proposals for *Sedimentitalea* gen. nov. and *Pseudophaeobacter* gen. nov." In: *Frontiers in Microbiology* 5.August (2014), pp. 1–13. DOI: 10.3389/fmicb.2014.00416.

[84] Marike Palmer, Emma T. Steenkamp, Martin P. A. Coetzee, Jochen Blom, and Stephanus N. Venter. "Genome-based characterization of biological processes that differentiate closely related bacteria." In: *Frontiers in Microbiology* 9.February (2018), pp. 1–23. DOI: 10.3389/fmicb.2018.00113.

[85] Ramon Rosselló-Móra and Rudolf Amann. "Past and future species definitions for bacteria and archaea." In: *Systematic and Applied Microbiology* 38.4 (2015), pp. 209–216. DOI: 10.1016/j.syapm.2015.02.001.

[86] Qi Long Qin, Bin Bin Xie, Xi Ying Zhang, Xiu Lan Chen, Bai Cheng Zhou, Jizhong Zhou, Aharon Oren, and Yu Zhong Zhang. "A proposed genus boundary for the prokaryotes based on genomic insights." In: *Journal of Bacteriology* 196.12 (2014), pp. 2210–2215. DOI: 10.1128/JB.01688-14.

[87] Pd Cantino and Kevin De Queiroz. *International code of phylogenetic nomenclature, Version 4c*. 2010. URL: https://www.ohio.edu/phylocode/PhyloCode4c.pdf{\%}0Ahttp://www.ohio.edu/PhyloCode/PhyloCode4c.pdf.

[88] Mark W. Williams, Sharon E. Stammerjohn, Hugh W. Ducklow, Andrew G. Fountain, Edward A. G. Schuur, and John L. Campbell. "The disappearing cryosphere: Impacts and ecosystem responses to rapid cryosphere loss." In: *BioScience* 62.4 (2012), pp. 405–415. DOI: 10.1525/bio.2012.62.4.11.

[89] Natascha Selje, Meinhard Simon, and Thorsten Brinkhoff. "A newly discovered Roseobacter cluster in temperate and polar oceans." In: *Nature* 427.6973 (2004), pp. 445–448. DOI: 10.1038/nature02272.

[90] Eileen Y. Koh, William Phua, and Ken G. Ryan. "Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater." In: *Environmental Microbiology Reports* 3.6 (2011), pp. 710–716. DOI: 10.1111/j.1758-2229.2011.00286.x.

[91] Frédéric Delsuc, Henner Brinkmann, and Hervé Philippe. "Phylogenomics and the reconstruction of the tree of life." In: *Nature Reviews Genetics* 6.5 (2005), pp. 361–375. DOI: 10.1038/nrg1603.

[92] Miklós Csurös and István Miklós. "Streamlining and large ancestral genomes in archaea inferred with a phylogenetic birth-and-death model." In: *Molecular Biology and Evolution* 26.9 (2009), pp. 2087–2095. DOI: 10.1093/molbev/msp123.

[93] Haiwei Luo, Miklós Csũros, Austin L Hughes, and Mary Ann Moran. "Evolution of divergent life history strategies in marine Alphaproteobacteria." In: *mBio* 4.4 (2013), e00373–13. DOI: 10.1128/mBio.00373-13.Editor.

[94] Riccardo Percudani, Davide Carnevali, and Vincenzo Puggioni. "Ureidoglycolate hydrolase, amidohydrolase, lyase: How errors in biological databases are incorporated in scientific papers and vice versa." In: *Database* 2013 (2013), bat071. DOI: 10.1093/database/bat071.

[95] Antoine Danchin. "Coping with inevitable accidents in metabolism." In: *Microbial Biotechnology* 10.1 (2017), pp. 57–72. DOI: 10.1111/1751-7915.12461.

[96] Antoine Danchin and Gang Fang. "Unknown unknowns: essential genes in quest for function." In: *Microbial Biotechnology* 9.5 (2016), pp. 530–540. DOI: 10.1111/1751-7915.12384.

[97] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. "Trimmomatic: A flexible trimmer for Illumina sequence data." In: *Bioinformatics* 30.15 (2014), pp. 2114–2120. DOI: 10.1093/bioinformatics/btu170.

[98] Anton Bankevich et al. "SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing." In: *Journal of Computational Biology* 19.5 (2012), pp. 455–477. DOI: 10.1089/cmb.2012.0021.

[99] Torsten Seemann. "Prokka: Rapid prokaryotic genome annotation." In: *Bioinformatics* 30.14 (2014), pp. 2068–2069. DOI: 10.1093/bioinformatics/btu153.

[100] Andreas Leimbach. *bac-genomics-scripts: Bovine E. coli mastitis comparative genomics edition*. 2016. DOI: 10.5281/zenodo.215824.

[101] Donovan H Parks, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and Gene W Tyson. "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." In: *Genome Research* 25 (2015), pp. 1043–1055. DOI: 10.1101/gr.186072.114.

[102] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. "KEGG: New perspectives on genomes, pathways, diseases and drugs." In: *Nucleic Acids Research* 45.D1 (2017), pp. D353–D361. DOI: 10.1093/nar/gkw1092.

[103] Jan P. Meier-Kolthoff, Alexander F. Auch, Hans Peter Klenk, and Markus Göker. "Genome sequence-based species delimitation with confidence intervals and improved distance functions." In: *BMC Bioinformatics* 14 (2013). DOI: 10.1186/1471-2105-14-60.

[104] Katherine J. Boettcher, Bruce J. Barber, and John T. Singer. "Use of antibacterial agents to elucidate the etiology of juvenile oyster disease (JOD) in *Crassostrea virginica* and numerical dominance of an α-Proteobacterium in JOD-affected animals." In: *Applied and Environmental Microbiology* 65.6 (1999), pp. 2534–2539.

[105] Qiliang Lai, Junwei Cao, Jun Yuan, Fuying Li, and Zongze Shao. "*Celeribacter indicus* sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium from deep-sea sediment and reclassification of *Huaishuia halophila* as *Celeribacter halophilus* comb. nov." In: *International Journal of Systematic and Evolutionary Microbiology* 64.2014 (2014), pp. 4160–4167. DOI: 10.1099/ijs.0.069039-0.

[106] L. Y. Suarez-Suarez, I. Brunet-Galmes, J. M. Piña Villalonga, J. A. Christie-Oleza, A. Peña, A. Bennasar, J. Armengaud, B. Nogales, and R. Bosch. "Draft genome sequence of *Citreicella aestuarii* strain 357, a member of the Roseobacter clade isolated without xenobiotic pressure from a petroleum-polluted beach." In: *Journal of Bacteriology* 194.19 (2012), pp. 5464–5465. DOI: 10.1128/JB.01261-12.

[107] Hanno Biebl, Martin Allgaier, Brian J. Tindall, Michal Kobližek, Heinrich Lünsdorf, Rüdiger Pukall, and Irene Wagner-Döbler. "*Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates." In: *Interna-*

tional Journal of Systematic and Evolutionary Microbiology 55.3 (2005), pp. 1089–1096. DOI: 10.1099/ijs.0.63511-0.

[108] Sooyeon Park and Jung Hoon Yoon. "*Jannaschia aquimarina* sp. nov., isolated from seawater." In: *International Journal of Systematic and Evolutionary Microbiology* 62.11 (2012), pp. 2631–2636. DOI: 10.1099/ijs.0.038448-0.

[109] M. C. Macián, D. R. Arahal, E. Garay, W. Ludwig, K. H. Schleifer, and Maria J. Pujalte. "*Jannaschia rubra* sp. nov., a red-pigmented bacterium isolated from sea water." In: *International Journal of Systematic and Evolutionary Microbiology* 55.2 (2005), pp. 649–653. DOI: 10.1099/ijs.0.63412-0.

[110] Liming Liu, Yang Li, Jing Zhang, Zhemin Zhou, Jie Liu, Xiaomin Li, Jingwen Zhou, Guocheng Du, Lei Wang, and Jian Chen. "Complete genome sequence of the industrial strain *Ketogulonicigenium vulgare* WSH-001." In: *Journal of Bacteriology* 193.21 (2011), pp. 6108–6109. DOI: 10.1128/JB.06007-11.

[111] Xiang Hua Xiong et al. "Complete genome sequence of the bacterium *Ketogulonicigenium vulgare* Y25." In: *Journal of Bacteriology* 193.1 (2011), pp. 315–316. DOI: 10.1128/JB.01189-10.

[112] Sooyeon Park, Yong Taek Jung, Su Jung Choi, and Jung Hoon Yoon. "*Phaeobacter aquaemixtae* sp. nov., isolated from the junction between the ocean and a freshwater spring." In: *International Journal of Systematic and Evolutionary Microbiology* 67.9 (2017), pp. 3446–3451. DOI: 10.1099/ijsem.0.002136.

[113] Ilse Vandecandelaere, Eveline Segaert, Alfonso Mollica, Marco Faimali, and Peter Vandamme. "*Leisingera aquimarina* sp. nov., isolated from a marine electroactive biofilm, and emended descriptions of *Leisingera methylohalidivorans* Schaefer et al. 2002, *Phaeobacter daeponensis* Yoon et al. 2007 and *Phaeobacter inhibens* Martens et al. 2006." In: *International Journal of Systematic and Evolutionary Microbiology* 58.12 (2008), pp. 2788–2793. DOI: 10.1099/ijs.0.65844-0.

[114] Ilse Vandecandelaere, Eveline Segaert, Alfonso Mollica, Marco Faimali, and Peter Vandamme. "*Phaeobacter caeruleus* sp. nov., a blue-coloured, colony-forming bacterium isolated from a marine electroactive biofilm." In: *International Journal of Systematic and Evolutionary Microbiology* 59.5 (2009), pp. 1209–1214. DOI: 10.1099/ijs.0.002642-0.

[115] Jung Hoon Yoon, So Jung Kang, Soo Young Lee, and Tae Kwang Oh. "*Phaeobacter daeponensis* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea." In: *International Journal of Systematic and Evolutionary Microbiology* 57.4 (2007), pp. 856–861. DOI: 10.1099/ijs.0.64779-0.

[116] Jk Schaefer, Kd Goodwin, Ir McDonald, J Colin Murrell, and Rs Oremland. "*Leisingera methylohalidivorans* gen. nov., sp. nov., a marine methylotroph that grows on methyl bromide." In: *International journal of systematic and evolutionary microbiology* 52.3 (2002), pp. 851–9. DOI: 10.1099/ijs.0.01960-0.The.

[117] Andrew J. Collins and Spencer V. Nyholm. "Draft genome of *Phaeobacter gallaeciensis* ANG1, a dominant member of the accessory nidamental gland of *Euprymna scolopes*." In: *Journal of Bacteriology* 193.13 (2011), pp. 3397–3398. DOI: 10.1128/JB.05139-11.

[118] Lyudmila A. Romanenko, Naoto Tanaka, Galina M. Frolova, Vassilii I. Svetashev, and Valery V. Mikhailov. "*Litoreibacter albidus* gen. nov., sp. nov. and *Litoreibacter janthinus* sp. nov., members of the class Alphaproteobacteria isolated from the seashore." In: *International Journal of Systematic and Evolutionary Microbiology* 61.1 (2011), pp. 148–154. DOI: 10.1099/ijs.0.019513-0.

[119] Byung Yong Kim, Hang Yeon Weon, Jung A. Son, Chang Muk Lee, Seung Beom Hong, Young Ah Jeon, Bon Sung Koo, and Soon Wo Kwon. "*Thalassobacter arenae* sp. nov., isolated from sea sand in Korea." In: *International Journal of Systematic and Evolutionary Microbiology* 59.3 (2009), pp. 487–490. DOI: 10.1099/ijs.0.65841-0.

[120] Young Ok Kim, In Suk Park, Sooyeon Park, Bo Hye Nam, Ji Min Park, Dong Gyun Kim, and Jung Hoon Yoon. "*Litoreibacter ascidiaceicola* sp. nov., isolated from the golden sea squirt *Halocynthia aurantium*." In: *International Journal of Systematic and Evolutionary Microbiology* 66.3 (2016), pp. 1174–1179. DOI: 10.1099/ijsem.0.000849.

[121] Shoichi Hosoya and Akira Yokota. "*Loktanella atrilutea* sp. nov., isolated from seawater in Japan." In: *International Journal of Systematic and Evolutionary Microbiology* 57.9 (2007), pp. 1966–1969. DOI: 10.1099/ijs.0.64978-0.

[122] Taishi Tsubouchi, Yasuhiro Shimane, Kozue Mori, Masayuki Miyazaki, Akihiro Tame, Katsuyuki Uematsu, Tadashi Maruyama, and Yuji Hatada. "*Loktanella cinnabarina* sp. nov., isolated from a deep subseafloor sediment, and emended description of the genus *Loktanella*." In: *International Journal of Systematic and Evolutionary Microbiology* 63.PART4 (2013), pp. 1390–1395. DOI: 10.1099/ijs.0.043174-0.

[123] Stanley C K Lau, Mandy M Y Tsoi, Xiancui Li, Ioulia Plakhotnikova, Madeline Wu, Po Keung Wong, and Pei Yuan Qian. "*Loktanella hongkongensis* sp. nov., a novel member of the α-Proteobacteria originating from marine biofilms in Hong Kong waters." In: *International Journal of Systematic and Evolutionary Microbiology* 54.6 (2004), pp. 2281–2284. DOI: 10.1099/ijs.0.63294-0.

[124] Hang Yeon Weon, Byung Yong Kim, Seung Hee Yoo, Jong Shik Kim, Soon Wo Kwon, Seung Joo Go, and Erko Stackebrandt. "*Loktanella koreensis* sp. nov., isolated from sea sand in Korea." In: *International Journal of Systematic and Evolutionary Microbiology* 56.9 (2006), pp. 2199–2202. DOI: 10.1099/ijs.0.64276-0.

[125] Jung Hoon Yoon, Yong Taek Jung, and Jung Sook Lee. "*Loktanella litorea* sp. nov., isolated from seawater." In: *International Journal of Systematic and Evolutionary Microbiology* 63.1 (2013), pp. 175–180. DOI: 10.1099/ijs.0.039198-0.

[126] Young Gun Moon, Seong Hae Seo, Soon Dong Lee, and Moon Soo Heo. "*Loktanella pyoseonensis* sp. nov., isolated from beach sand, and emended description of the genus *Loktanella*." In: *International Journal of Systematic and Evolutionary Microbiology* 60.4 (2010), pp. 785–789. DOI: 10.1099/ijs.0.011072-0.

[127] Elena P. Ivanova, Natalia V. Zhukova, Anatoly M. Lysenko, Nataliya M. Gorshkova, Alexander F. Sergeev, Valery V. Mikhailov, and John P. Bowman. "*Loktanella agnita* sp. nov. and *Loktanella rosea* sp. nov., from the north-west Pacific Ocean." In: *International Journal of Systematic and Evolutionary Microbiology* 55.5 (2005), pp. 2203–2207. DOI: 10.1099/ijs.0.63461-0.

[128] J. Liang, Z. Zhang, Y. Liu, M. Wang, and X.-H. Zhang. "*Loktanella sediminum* sp. nov., isolated from marine surface sediment." In: *International Journal of Systematic and Evolutionary Microbiology* 65.Pt 2 (2015), pp. 686–691. DOI: 10.1099/ijs.0.070680-0.

[129] Soon Dong Lee. "*Loktanella tamlensis* sp. nov., isolated from seawater." In: *International Journal of Systematic and Evolutionary Microbiology* 62.PART 3 (2012), pp. 586–590. DOI: 10.1099/ijs.0.029462-0.

[130] David H. Green, Lyndon E. Llewellyn, Andrew P. Negri, Susan I. Blackburn, and Christopher J.S. Bolch. "Phylogenetic and functional diversity of the cultivable bacterial community associated with the paralytic shellfish poisoning dinoflagellate *Gymnodinium catenatum*." In: *FEMS Microbiology Ecology* 47.3 (2004), pp. 345–357. DOI: 10.1016/S0168-6496(03)00298-8.

[131] Kiyoung Lee, Yoe Jin Choo, Stephen J. Giovannoni, and Jang Cheon Cho. "*Maritimibacter alkaliphilus* gen. nov., sp. nov., a genome-sequenced marine bacterium of the Roseobacter clade in the order *Rhodobacterales*." In: *International Journal of Systematic and Evolutionary Microbiology* 57.7 (2007), pp. 1653–1658. DOI: 10.1099/ijs.0.64960-0.

[132] Ilse Vandecandelaere, Olivier Nercessian, Eveline Segaert, Wafa Achouak, Alfonso Mollica, Marco Faimali, and Peter Vandamme. "*Nautella italica* gen. nov., sp. nov., isolated from a marine electroactive biofilm." In: *International Journal of Systematic and Evolutionary Microbiology* 59.4 (2009), pp. 811–817. DOI: 10.1099/ijs.0.002683-0.

[133] Maria J. Pujalte, M. C. Macián, D. R. Arahal, W. Ludwig, K. H. Schleifer, and E. Garay. "*Nereida ignava* gen. nov., sp. nov., a novel aerobic marine α-Proteobacterium that is closely related to uncultured *Prionitis* (alga) gall symbionts." In: *International Journal of Systematic and Evolutionary Microbiology* 55.2 (2005), pp. 631–636. DOI: 10.1099/ijs.0.63442-0.

[134] Jang Cheon Cho and Stephen J. Giovannoni. "*Oceanicola granulosus* gen. nov., sp. nov. and *Oceanicola batsensis* sp. nov., poly-β-hydroxybutyrate-producing marine bacteria in the order *'Rhodobacterales'*." In: *International Journal of Systematic and Evolutionary Microbiology* 54.4 (2004), pp. 1129–1136. DOI: 10.1099/ijs.0.03015-0.

[135] Jun Gu, Bin Guo, Ya-Nan Wang, Su-Lin Yu, Ryuhei Inamori, Ri Qu, Yu-Guang Ye, and Xiao-Lei Wu. "*Oceanicola nanhaiensis* sp. nov., isolated from sediments of the South China Sea." In: *International journal of systematic and evolutionary microbiology* 57.2007 (2007), pp. 157–160. DOI: 10.1099/ijs.0.64532-0.

[136] Tony Gutierrez, William B Whitman, Marcel Huntemann, Alex Copeland, Amy Chen, Neha Vargese, Nikos C Kyrpides, Manoj Pillay, and Natalia Ivanova. "Genome sequence of *Oceanicola* sp. Strain MCTG156(1a), isolated from a Scottish coastal phytoplankton net sample." In: *Genome Announcements* 5.32 (2017), pp. 9–10. DOI: 10.1128/genomeA.00796-17.

[137] Young Kyung Kwon, Jennifer Jooyoun Kim, Ji Hyung Kim, Seon Mi Jeon, Bo Ram Ye, Jiyi Jang, Soo Jin Heo, Se Chang Park, Do Hyung Kang, and Chulhong Oh. "Draft genome sequence of the xylan-degrading marine bacterium strain S124, representing a novel species of the genus *Oceanicola*." In: *Journal of Bacteriology* 194.22 (2012), p. 6325. DOI: 10.1128/JB.01614-12.

[138] Keshao Liu, Rui Zong, Qipei Li, Yingnan Fu, Yongle Xu, Yanan Wang, and Nianzhi Jiao. "*Oceaniovalibus guishaninsula* gen. nov., sp. nov., a marine bacterium of the family *Rhodobacteraceae*." In: *Current Microbiology* 64.4 (2012), pp. 385–391. DOI: 10.1007/s00284-012-0081-z.

[139] Jang Cheon Cho and Stephen J. Giovannoni. "*Parvularcula bermudensis* gen. nov., sp. nov., a marine bacterium that forms a deep branch in the α-Proteobacteria." In: *International Journal of Systematic and Evolutionary Microbiology* 53.4 (2003), pp. 1031–1036. DOI: 10.1099/ijs.0.02566-0.

[140] Jang Cheon Cho and Stephen J. Giovannoni. "*Pelagibaca bermudensis* gen. nov., sp. nov., a novel marine bacterium within the Roseobacter clade in the order *Rhodobacterales*." In: *International Journal of Systematic and Evolutionary Microbiology* 56.4 (2006), pp. 855–859. DOI: 10.1099/ijs.0.64063-0.

[141] C Ruiz-Ponte, V Cilia, C Lambert, and J L Nicolas. "*Roseobacter gallaeciensis* sp. nov., a new marine bacterium isolated from rearings and collectors of the scallop *Pecten maximus*." In: *International Journal of Systematic Bacteriology* 48 (1998), pp. 537–542.

[142] Torben Martens, Thorsten Heidorn, Rüdiger Pukal, Meinhard Simon, Brian J. Tindall, and Thorsten Brinkhoff. "Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., reclassification of *Ruegeria algicola* (Lafay et al. 1995) Uchino et al. 1999 as *Marinovum algicola* gen. nov., comb. nov., and emended descriptions of the genera *Roseobacter*, *Ruegeria* and *Leisingera*." In: *International Journal of Systematic and Evolutionary Microbiology* 56.6 (2006), pp. 1293–1304. DOI: 10.1099/ijs.0.63724-0.

[143] Nora Buddruhs, Silke Pradella, Markus Goker, Orsola Pauker, Rüdiger Pukall, Cathrin Sproer, Peter Schumann, Jörn Petersen, and Thorsten Brinkhoff. "Molecular and phenotypic analyses reveal the non-identity of the *Phaeobacter gallaeciensis* type strain deposits CIP 105210T and DSM 17395." In: *International Journal of Systematic and Evolutionary Microbiology* 63.PART 11 (2013), pp. 4340–4349. DOI: 10.1099/ijs.0.053900-0.

[144] Helge Ansgar Giebel et al. "*Planktomarina temperata* gen. nov., sp. nov., belonging to the globally distributed RCA cluster of the marine Roseobacter clade, isolated from the German Wadden Sea." In: *International Journal of Systematic and Evolutionary Microbiology* 63.PART 11 (2013), pp. 4207–4217. DOI: 10.1099/ijs.0.053249-0.

[145] Ilnam Kang, Kevin L. Vergin, Hyun Myung Oh, Ahyoung Choi, Stephen J. Giovannoni, and Jang Cheon Cho. "Genome sequence of strain HTCC2083, a novel member of the marine clade Roseobacter." In: *Journal of Bacteriology* 193.1 (2011), pp. 319–320. DOI: 10.1128/JB.01268-10.

[146] Ilnam Kang, Hyun Myung Oh, Kevin L. Vergin, Stephen J. Giovannoni, and Jang Cheon Cho. "Genome sequence of the marine alphaproteobacterium HTCC2150, assigned to the Roseobacter clade." In: *Journal of Bacteriology* 192.23 (2010), pp. 6315–6316. DOI: 10.1128/JB.01088-10.

[147] Colleen M Hansel and Chris A Francis. "Coupled photochemical and enzymatic Mn(II) oxidation pathways of a planktonic Roseobacter-like bacterium." In: *Applied and Environmental Microbiology* 72.Ii (2006), pp. 3543–3549. DOI: 10.1128/AEM.72.5.3543.

[148] Guizhen Li, Qiliang Lai, Chunming Dong, Rui Ma, Yaping Du, Xiupian Liu, Fengqin Sun, Zongze Shao, and Peisheng Yan. "*Roseovarius atlanticus* sp. nov., isolated from surface seawater."

In: *International Journal of Systematic and Evolutionary Microbiology* 66.2 (2016), pp. 639–644. DOI: 10.1099/ijsem.0.000768.

[149] Hanno Biebl, Martin Allgaier, Heinrich Lünsdorf, Rüdiger Pukall, Brian J. Tindall, and Irene Wagner-Döbler. "*Roseovarius mucosus* sp. nov., a member of the Roseobacter clade with trace amounts of bacteriochlorophyll a." In: *International Journal of Systematic and Evolutionary Microbiology* 55.6 (2005), pp. 2377–2383. DOI: 10.1099/ijs.0.63832-0.

[150] José M. González et al. "*Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments." In: *International Journal of Systematic and Evolutionary Microbiology* 53.5 (2003), pp. 1261–1269. DOI: 10.1099/ijs.0.02491-0.

[151] Hendrik Schäfer, Ian R. McDonald, Phil D. Nightingale, and J. Colin Murrell. "Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria." In: *Environmental Microbiology* 7.6 (2005), pp. 839–852. DOI: 10.1111/j.1462-2920.2005.00757.x.

[152] Tony Gutierrez et al. "Genome sequence of *Roseovarius* sp. strain MCTG156(2b) isolated from a phytoplankton net trawl on the Scottish West Coast." In: *Genome Announcements* 5.35 (2017), pp. 1–2.

[153] Hans-Jürgen Rüger and Manfred G Höfle. "Marine star-shaped-aggregate-forming bacteria : *Agrobacterium atlanticum* sp. nov.; *Agrobacterium meteori* sp. nov.; *Agrobacterium ferrugineum* sp. nov., nom. rev.; *Agrobacterium gelatinovorum* sp. nov., nom. rev.; and *Agrobacterium stellulatum* sp. nov., nom. rev." In: *International Journal of Systematic Bacteriology* 42.1 (1992), pp. 133–143. DOI: 10.1099/00207713-42-1-133.

[154] Jina Lee et al. "*Ruegeria conchae* sp. nov., isolated from the ark clam *Scapharca broughtonii*." In: *International Journal of Systematic and Evolutionary Microbiology* 62.12 (2012), pp. 2851–2857. DOI: 10.1099/ijs.0.037283-0.

[155] Margot Doberva, Sophie Sanchez-Ferandin, Yoan Ferandin, Laurent Intertaglia, Julie Croué, Marcelino Suzuki, Philippe Lebaron, and Raphaël Lami. "Genome sequence of the sponge-associated *Ruegeria halocynthiae* strain MOLA R1/13b, a marine Roseobacter with two quorum-sensing-based communication systems." In: *Genome announcements* 2.5 (2014), p. 4389. DOI: 10.1128/genomeA.00998-14.

[156] Eva C. Sonnenschein et al. "Global occurrence and heterogeneity of the Roseobacter-clade species *Ruegeria mobilis*." In: *ISME Journal* 11.2 (2017), pp. 569–583. DOI: 10.1038/ismej.2016.111.

[157] J M Gonzalez, F Mayer, M A Moran, and R E Hodson. "*Sagittula stellata* gen. nov., sp. nov., a lignin-transforming bacterium from a coastal environment." In: *International Journal of Systematic Bacteriology* 47.3 (1997), pp. 773–780. DOI: 10.1099/00207713-47-3-773.

[158] Fengqin Sun, Baojiang Wang, Xiupian Liu, Qiliang Lai, Yaping Du, Guangyu Li, Jie Luo, and Zongze Shao. "*Leisingera nanhaiensis* sp. nov., isolated from marine sediment." In: *International Journal of Systematic and Evolutionary Microbiology* 60.2 (2010), pp. 275–280. DOI: 10.1099/ijs.0.010439-0.

[159] Dong H. Choi and Byung C. Cho. "*Shimia marina* gen. nov., sp. nov., a novel bacterium of the Roseobacter clade isolated from biofilm in a coastal fish farm." In: *International Journal of Systematic and Evolutionary Microbiology* 56.8 (2006), pp. 1869–1873. DOI: 10.1099/ijs.0.64235-0.

[160] Saranya Kanukollu et al. "Genome sequence of *Shimia* str. SK013, a representative of the *Roseobacter* group isolated from marine sediment." In: *Standards in Genomic Sciences* 11.1 (2016). DOI: 10.1186/s40793-016-0143-0.

[161] Jung Hoon Yoon, So Jung Kang, Mi Hwa Lee, and Tae Kwang Oh. "Description of *Sulfitobacter donghicola* sp. nov., isolated from seawater of the East Sea in Korea, transfer of *Staleya guttiformis* Labrenz et al. 2000 to the genus *Sulfitobacter* as *Sulfitobacter guttiformis* comb. nov. and emended description of the genus *Sulfitobacter*." In: *International Journal of Systematic and Evolutionary Microbiology* 57.8 (2007), pp. 1788–1792. DOI: 10.1099/ijs.0.65071-0.

[162] Min Jung Kwak, Jung Sook Lee, Keun Chul Lee, Kwang Kyu Kim, Mi Kyung Eom, Byung Kwon Kim, and Jihyun F. Kim. "*Sulfitobacter geojensis* sp. nov., *Sulfitobacter noctilucae* sp. nov., and *Sulfitobacter noctilucicola* sp. nov., isolated from coastal seawater." In: *International Journal of Systematic and Evolutionary Microbiology* 64.2014 (2014), pp. 3760–3767. DOI: 10.1099/ijs.0.065961-0.

[163] Matthias Labrenz, B. J. Tindall, Paul A. Lawson, Matthew D. Collins, Peter Schumann, and Peter Hirsch. "*Staleya guttiformis* gen. nov., sp. nov. and *Sulfitobacter brevis* sp. nov., α-3-Proteobacteria from hypersaline, heliothermal and meromictic antarctic Ekho Lake." In: *International Journal of Systematic and Evolutionary Microbiology* 50.1 (2000), pp. 303–313. DOI: 10.1099/00207713-50-1-303.

[164] Rüdiger Pukall, Daniela Buntefuß, Anja Frühling, Manfred Rohde, Reiner M Kroppenstedt, Jutta Burghardt, Philippe Lebar, Laetitia Bernard, and Erko Stackebrandtl. "*Sulfitobacter mediterraneus* sp. nov., a new sulfite-oxidizing member of the α-Proteobacteria." In: *International Journal of Systematic Bacteriology* 49.2 (1999), pp. 513–519. DOI: 10.1099/00207713-49-2-513.

[165] Maria Mas-lladó, Joana Maria Pi na-villalonga, Isabel Brunet-galmés, and Balbina Nogales. "Draft genome sequences of two isolates of the *Roseobacter* group, *Sulfitobacter* sp. Strains 3SOLIMAR09 and 1FIGIMAR09, from harbors of Mallorca Island (Mediterranean Sea)." In: *Genome Announcements* 2.3 (2014), e00350–14. DOI: 10.1128/genomeA.00350-14.Copyright.

[166] Zhuan Hong, Qiliang Lai, Qiaoqi Luo, Simeng Jiang, Ruilin Zhu, Junrong Liang, and Yahui Gao. "*Sulfitobacter pseudonitzschiae* sp. nov., isolated from the toxic marine diatom *Pseudo-nitzschia multiseries*." In: *International Journal of Systematic and Evolutionary Microbiology* 65.1 (2015), pp. 95–100. DOI: 10.1099/ijs.0.064972-0.

[167] N. Y. D. Ankrah, T. Lane, C. R. Budinoff, M. K. Hadden, and A. Buchan. "Draft genome sequence of *Sulfitobacter* sp. CB2047, a member of the Roseobacter clade of marine bacteria, isolated from an *Emiliania huxleyi* bloom." In: *Genome Announcements* 2.6 (2014), e01125–14–e01125–14. DOI: 10.1128/genomeA.01125-14.

[168] Andrew J. Collins, Matthew S. Fullmer, Johann P. Gogarten, and Spencer V. Nyholm. "Comparative genomics of Roseobacter clade bacteria isolated from the accessory nidamental gland of *Euprymna scolopes*." In: *Frontiers in Microbiology* 6.FEB (2015), pp. 1–14. DOI: 10.3389/fmicb.2015.00123.

[169] M. C. Macián, D. R. Arahal, E. Garay, W. Ludwig, K. H. Schleifer, and Maria J. Pujalte. "*Thalassobacter stenotrophicus* gen. nov., sp. nov., a novel marine α-Proteobacterium isolated from Mediterranean sea water." In: *International Journal of Systematic and Evolutionary Microbiology* 55.1 (2005), pp. 105–110. DOI: 10.1099/ijs.0.63275-0.

[170] Jiao Yan Ying, Bao Jun Wang, Xin Dai, Su Sheng Yang, Shuang Jiang Liu, and Zhi Pei Liu. "*Wenxinia marina* gen. nov., sp. nov., a novel member of the Roseobacter clade isolated from oilfield sediments of the South China Sea." In: *International Journal of Systematic and Evolutionary Microbiology* 57.8 (2007), pp. 1711–1716. DOI: 10.1099/ijs.0.64825-0.

[171] Marcus Lechner, Sven Findeiß, Lydia Steiner, Manja Marz, Peter F Stadler, and Sonja J Prohaska. "Proteinortho : detection of (co-)orthologs in large-scale analysis." In: *BMC Bioinformatics* 12.1 (2011), p. 124. DOI: 10.1186/1471-2105-12-124.

[172] Robert C. Edgar. "MUSCLE: Multiple sequence alignment with high accuracy and high throughput." In: *Nucleic Acids Research* 32.5 (2004), pp. 1792–1797. DOI: 10.1093/nar/gkh340.

[173] Alexandros Stamatakis. "RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies." In: *Bioinformatics* 30.9 (2014), pp. 1312–1313. DOI: 10.1093/bioinformatics/btu033.

[174] Leonidas Salichos, Alexandros Stamatakis, and Antonis Rokas. "Novel information theory-based measures for quantifying incongruence among phylogenetic trees." In: *Molecular Biology and Evolution* 31.5 (2014), pp. 1261–1271. DOI: 10.1093/molbev/msu061.

[175] F. Pedregosa et al. "Scikit-learn: machine learning in Python." In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[176] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. [Online; accessed <today>]. 2001– . URL: http://www.scipy.org/.

[177] Jaime Huerta-Cepas, François Serra, and Peer Bork. "ETE 3: Reconstruction, analysis, and visualization of phylogenomic data." In: *Molecular Biology and Evolution* 33.6 (2016), pp. 1635–1638. DOI: 10.1093/molbev/msw046.

[178] Daniel H. Huson and Celine Scornavacca. "Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks." In: *Systematic Biology* 61.6 (2012), pp. 1061–1067. DOI: 10.1093/sysbio/sys062.

[179] Natalia N Ivanova, Konstantinos Mavromatis, I-min a Chen, Victor M Markowitz, and Nikos C Kyrpides. *Standard operating procedure for the annotations of genomes and metagenomes submitted to the Integrated Microbial Genomes Expert Review (IMG-ER) system*. 2007. URL: http://imgweb.jgi-psf.org/w/doc/img_er_ann.pdf.

[180] Michael Y. Galperin, Kira S. Makarova, Yuri I. Wolf, and Eugene V. Koonin. "Expanded microbial genome coverage and improved protein family annotation in the COG database." In: *Nucleic Acids Research* 43.D1 (2015), pp. D261–D269. DOI: 10.1093/nar/gku1223.

[181] Miklós Csurös. "Count: Evolutionary analysis of phylogenetic profiles with parsimony and likelihood." In: *Bioinformatics* 26.15 (2010), pp. 1910–1912. DOI: 10.1093/bioinformatics/btq315.

[182] Alexey Drozdetskiy, Christian Cole, James Procter, and Geoffrey J. Barton. "JPred4: A protein secondary structure prediction server." In: *Nucleic Acids Research* 43.W1 (2015), W389–W394. DOI: 10.1093/nar/gkv332.

[183] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1 (1995), pp. 289–300. DOI: 10.2307/2346101.

[184] Folker Meyer, D Paarmann, M D'Souza, and Etal. "The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes." In: *BMC bioinformatics* 9 (2008), p. 386. DOI: 10.1186/1471-2105-9-386.

[185] Karin Lagesen, Peter Hallin, Einar Andreas Rødland, Hans Henrik Stærfeldt, Torbjørn Rognes, and David W. Ussery. "RNAmmer: Consistent and rapid annotation of ribosomal RNA genes." In: *Nucleic Acids Research* 35.9 (2007), pp. 3100–3108. DOI: 10.1093/nar/gkm160.

[186] Ben Langmead and Steven L. Salzberg. "Fast gapped-read alignment with Bowtie 2." In: *Nature Methods* 9.4 (2012), pp. 357–359. DOI: 10.1038/nmeth.1923.

[187] Leighton Pritchard, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. "Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens." In: *Analytical Methods* 8.1 (2016), pp. 12–24. DOI: 10.1039/c5ay02550h.

[188] George Vernikos, Duccio Medini, David R. Riley, and Hervé Tettelin. "Ten years of pan-genome analyses." In: *Current Opinion in Microbiology* 23 (2015), pp. 148–154. DOI: 10.1016/j.mib.2014.11.016.

[189] Eugene V. Koonin and Yuri I. Wolf. "Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world." In: *Nucleic Acids Research* 36.21 (2008), pp. 6688–6719. DOI: 10.1093/nar/gkn668.

[190] R. Overbeek, M. Fonstein, M. D'Souza, G. D. Pusch, and N. Maltsev. "The use of gene clusters to infer functional coupling." In: *Proceedings of the National Academy of Sciences* 96.6 (1999), pp. 2896–2901. DOI: 10.1073/pnas.96.6.2896.

[191] Inoka C. Perera and Anne Grove. "Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators." In: *Journal of Molecular Cell Biology* 2.5 (2010), pp. 243–254. DOI: 10.1093/jmcb/mjq021.

[192] Ziyang Hao et al. "The multiple antibiotic resistance regulator MarR is a copper sensor in *Escherichia coli*." In: *Nature Chemical Biology* 10.1 (2014), pp. 21–28. DOI: 10.1038/nchembio.1380.

[193] Steven P. Wilkinson and Anne Grove. "Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins." In: *Current Issues in Molecular Biology* 8.1 (2006), pp. 51–62.

[194] Béatrice Py et al. "Molecular organization, biochemical function, cellular role and evolution of NfuA, an atypical Fe-S carrier." In: *Molecular Microbiology* 86.1 (2012), pp. 155–171. DOI: 10.1111/j.1365-2958.2012.08181.x.

[195] Kai Zhao, Mingzhu Liu, and Richard R. Burgess. "The global transcriptional response of *Escherichia coli* to induced σ32 protein involves σ32 regulon activation followed by inactivation and degradation of σ32 in vivo." In: *Journal of Biological Chemistry* 280.18 (2005), pp. 17758–17768. DOI: 10.1074/jbc.M500393200.

[196] Phyllis I. Hanson and Sidney W. Whiteheart. "AAA+ proteins: Have engine, will work." In: *Nature Reviews Molecular Cell Biology* 6.7 (2005), pp. 519–529. DOI: 10.1038/nrm1684.

[197] Ryuta Tobe and Hisaaki Mihara. "Delivery of selenium to selenophosphate synthetase for selenoprotein biosynthesis." In: *Biochimica et Biophysica Acta - General Subjects* 1862.11 (2018), pp. 2433–2440. DOI: 10.1016/j.bbagen.2018.05.023.

[198] Agathe Tarze, Marc Dauplais, Ioana Grigoras, Myriam Lazard, Nguyet Thanh Ha-Duong, Frédérique Barbier, Sylvain Blanquet, and Pierre Plateau. "Extracellular production of hydrogen selenide accounts for thiol-assisted toxicity of selenite against *Saccharomyces cerevisiae*." In: *Journal of Biological Chemistry* 282.12 (2007), pp. 8759–8767. DOI: 10.1074/jbc.M610078200.

[199] Y. Ogasawara, G. M. Lacourciere, K. Ishii, and T. C. Stadtman. "Characterization of potential selenium-binding proteins in the selenophosphate synthetase system." In: *Proceedings of the National Academy of Sciences* 102.4 (2005), pp. 1012–1016. DOI: 10.1073/pnas.0409042102.

[200] Obed Béjà et al. "Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea." In: *Science* 289.5486 (2000), pp. 1902–1906. DOI: 10.1126/science.289.5486.1902.

[201] Omri M Finkel, Oded Béjà, and Shimshon Belkin. "Global abundance of microbial rhodopsins." In: *The ISME journal* 7.2 (2013), pp. 448–51. DOI: 10.1038/ismej.2012.112.

[202] Jed a Fuhrman, Michael S Schwalbach, and Ulrich Stingl. "Proteorhodopsins: An array of physiological roles?" In: *Nature reviews. Microbiology* 6.6 (2008), pp. 488–94. DOI: 10.1038/nrmicro1893.

[203] Sergei P Balashov, Eleonora S Imasheva, Vladimir a Boichenko, Josefa Antón, Jennifer M Wang, and Janos K Lanyi. "Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna." In: *Science* 309.5743 (2005), pp. 2061–4. DOI: 10.1126/science.1118046.

[204] J. M. Walter, D. Greenfield, C. Bustamante, and J. Liphardt. "Light-powering *Escherichia coli* with proteorhodopsin." In: *Proceedings of the National Academy of Sciences* 104.7 (2007), pp. 2408–2412. DOI: 10.1073/pnas.0611035104.

[205] Laura Gómez-Consarnau, Neelam Akram, Kristoffer Lindell, Anders Pedersen, Richard Neutze, Debra L. Milton, José M. González, and Jarone Pinhassi. "Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation." In: *PLoS Biology* 8.4 (2010), pp. 2–11. DOI: 10.1371/journal.pbio.1000358.

[206] Michael J Weickert and Sankar Adhya. "A family of bacterial regulators homologous to Gal and Lac repressors." In: *Journal of Biological Chemistry* 267.22 (1992), pp. 15869–15874.

[207] Erwin Schneider. "ABC transporters catalyzing carbohydrate uptake." In: *Research in Microbiology* 152.3-4 (2001), pp. 303–310. DOI: 10.1016/S0923-2508(01)01201-3.

[208] E C C Lin. "Glycerol dissimilation and its regulation in bacteria." In: *Annual Review of Microbiology* 30.1 (1976), pp. 535–578. DOI: 10.1146/annurev.mi.30.100176.002535.

[209] Rahel Elevi Bardavid and Aharon Oren. "Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Haloquadratum walsbyi*." In: *Extremophiles* 12.1 (2008), pp. 125–131. DOI: 10.1007/s00792-007-0114-x.

[210] K I Sørensen and B Hove-Jensen. "Ribose catabolism of *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and of the *rpiR* gene, which is involved in regulation of *rpiB* expression." In: *Journal of bacteriology* 178.4 (1996), pp. 1003–11. DOI: 10.1016/j.jfca.2014.04.004.

[211] Yefei Zhu, Renu Nandakumar, Marat R. Sadykov, Nandakumar Madayiputhiya, Thanh T. Luong, Rosmarie Gaupp, Chia Y. Lee, and Greg A. Somerville. "RpiR homologues may link *Staphylococcus aureus* RNAIII synthesis and pentose phosphate pathway regulation." In: *Journal of Bacteriology* 193.22 (2011), pp. 6187–6196. DOI: 10.1128/JB.05930-11.

[212] Abdelali Daddaoua, Tino Krell, and Juan Luis Ramos. "Regulation of glucose metabolism in *Pseudomonas*. The phosphorylative branch and Entner-Doudoroff enzymes are regulated by a repressor containing a sugar isomerase domain." In: *Journal of Biological Chemistry* 284.32 (2009), pp. 21360–21368. DOI: 10.1074/jbc.M109.014555.

[213] Christopher Mulligan, Marcus Fischer, and Gavin H. Thomas. "Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea." In: *FEMS Microbiology Reviews* 35.1 (2011), pp. 68–86. DOI: 10.1111/j.1574-6976.2010.00236.x.

[214] Gavin H. Thomas, Thomas Southworth, Maria Rocio León-Kempis, and Andrew Leech. "Novel ligands for the extracellular solute receptors of two bacterial TRAP transporters." In: *Microbiology* 152.1 (2006), pp. 187–198. DOI: 10.1099/mic.0.28334-0.

[215] Shin Kurihara, Shinpei Oda, Kenji Kato, Hyeon Guk Kim, Takashi Koyanagi, Hidehiko Kumagai, and Hideyuki Suzuki. "A novel putrescine utilization pathway involves γ-glutamylated intermediates of *Escherichia coli* K-12." In: *Journal of Biological Chemistry* 280.6 (2005), pp. 4602–4608. DOI: 10.1074/jbc.M411114200.

[216] H. J. Rhee, Eui Jin Kim, and J. K. Lee. "Physiological polyamines: Simple primordial stress molecules." In: *Journal of Cellular and Molecular Medicine* 11.4 (2007), pp. 685–703. DOI: 10.1111/j.1582-4934.2007.00077.x.

[217] John A. Leigh and Jeremy A. Dodsworth. "Nitrogen regulation in bacteria and archaea." In: *Annual Review of Microbiology* 61.1 (2007), pp. 349–377. DOI: 10.1146/annurev.micro.61.080706.093409.

[218] Juan L Ramos, Manuel Martínez-Bueno, Antonio J Molina-Henares, Wilson Terán, Kazuya Watanabe, Xiaodong Zhang, María Trinidad Gallegos, Richard Brennan, and Raquel Tobes. "The TetR family of transcriptional repressors." In: *Microbiology and Molecular Biology Reviews* 69.2 (2005), pp. 326–356. DOI: 10.1128/MMBR.69.2.326-356.2005.

[219] Tatiana Claro da Silva, James E Polli, and Peter W Swaan. "The solute carrier family 10 (SLC10): beyond bile acid transport." In: *Molecular Aspects of Medicine* 34.2-3 (2013), pp. 252–269. DOI: 10.1016/j.mam.2012.07.004.

[220] Nien Jen Hu, So Iwata, Alexander D. Cameron, and David Drew. "Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT." In: *Nature* 478.7369 (2011), pp. 408–411. DOI: 10.1038/nature10450.

[221] Bodo Philipp. "Bacterial degradation of bile salts." In: *Applied Microbiology and Biotechnology* 89.4 (2011), pp. 903–915. DOI: 10.1007/s00253-010-2998-0.

[222] W. A. Anderson and Boris Magasanik. "The pathway of *myo*-inositol degradation in *Aerobacter aerogenes: Identification of the intermediate 2-deoxy-5-keto-D-gluconic acid*." In: *Biological Chemistry* 246.18 (1971), pp. 5658–5661.

[223] W A Anderson and Boris Magasanik. "The pathway of *myo*-inositol degradation in *Aerobacter aerogenes*: Conversion of 2-deoxy-5-keto-D-gluconic acid to glycolytic intermediates." In: *Biological Chemistry* 246.18 (1971), pp. 5662–5675.

[224] Thomas Berman and Boris Magasanik. "The pathway of *myo*-inositol degradation in *Aerobacter aerogenes*: Dehydrogenation and dehydration." In: *Biological Chemistry* 241.4 (1966), pp. 800–806.

[225] Thomas Berman and Boris Magasanik. "The pathway of *myo*-inositol degradation in *Aerobacter aerogenes*: Ring scission." In: *Biological Chemistry* 241.4 (1966), pp. 807–813.

[226] Ken Ichi Yoshida, Masanori Yamaguchi, Tetsuro Morinaga, Masaki Kinehara, Maya Ikeuchi, Hitoshi Ashida, and Yasutaro Fujita. "*myo*-inositol catabolism in *Bacillus subtilis*." In: *Journal of Biological Chemistry* 283.16 (2008), pp. 10415–10424. DOI: 10.1074/jbc.M708043200.

[227] Mohammad Ali Ghalambor and Edward C Heath. "The metabolism of L-Fucose." In: *Journal of Biological Chemistry* 237.8 (1962), pp. 2423–2427.

[228] Myron F. Goodman and Roger Woodgate. "Translesion DNA polymerases." In: *Cold Spring Harbor Perspectives in Biology* 5.10 (2013), pp. 1–20. DOI: 10.1101/cshperspect.a010363.

[229] Charles R Booth and Sasha Madronich. "Radiation amplification factors: Improved formulation accounts for large increases in ultraviolet radiation associated with Antarctic ozone depletion." In: *Ultraviolet Radiation in Antarctica: Measurements and Biological Effects*. Ed. by Susan C Weiler and Polly A Penhale. Washington, D.C.: American Geophysical Union, 1994. Chap. 3, pp. 39–42. ISBN: 9780875908410. DOI: 10.1029/AR062.

[230] B. Yeiser, E. D. Pepper, M. F. Goodman, and S. E. Finkel. "SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness." In: *Proceedings of the National Academy of Sciences* 99.13 (2002), pp. 8737–8741. DOI: 10.1073/pnas.092269199.

[231] Zijun Xiao and Ping Xu. "Acetoin metabolism in bacteria." In: *Critical Reviews in Microbiology* 33.2 (2007), pp. 127–140. DOI: 10.1080/10408410701364604.

[232] David White, James Drummond, and Clay Fuqua. *The physiology and biochemistry of prokaryotes*. 4. New York: Oxford University Press, 2012. ISBN: 978-0-19-539304-0.

[233] E Almaas, B Kovács, T Vicsek, ZN Oltvai, and A-L Barabási. "Global organization of metabolic fluxes in the bacterium *Escherichia coli*." In: *Nature* 427.6977 (2004), pp. 839–43. DOI: 10.1038/nature02289.

[234] Sébastien Zappa, Keran Li, and Carl E Bauer. "The tetrapyrrole biosynthetic pathway and its regulation in *Rhodobacter capsulatus*." In: *Recent Advances in Phototrophic Prokaryotes*. Ed. by Patrick C Hallenbeck. 1. Springer, 2010. Chap. 13, pp. 229–250. ISBN: 978-1-4419-1527-6. DOI: 10.1007/978-1-4419-1528-3.

[235] H. Mihara and N. Esaki. "Bacterial cysteine desulfurases: Their function and mechanisms." In: *Applied Microbiology and Biotechnology* 60.1-2 (2002), pp. 12–23. DOI: 10.1007/s00253-002-1107-4.

[236] Ron Ortenberg and Jon Beckwith. "Functions of thiol-disulfide oxidoreductases in *E. coli*: Redox myths, realities, and practicalities." In: *Antioxidants & Redox Signaling* 5.4 (2003), pp. 403–411. DOI: 10.1089/15230860376829514O.

[237] Krzysztof Ginalski, Lisa Kinch, Leszek Rychlewski, and Nick V. Grishin. "DCC proteins: a novel family of thiol-disulfide oxidoreductases." In: *Trends in Biochemical Sciences* 29.7 (2004), pp. 339–342. DOI: 10.1016/j.tibs.2004.05.002.

[238] Eric Eichhorn, Jan R. Van Der Ploeg, Michael A. Kertesz, and Thomas Leisinger. "Characterization of α-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*." In: *Journal of Biological Chemistry* 272.37 (1997), pp. 23031–23036. DOI: 10.1074/jbc.272.37.23031.

[239] L. Pollegioni, L. Piubelli, S. Sacchi, M. S. Pilone, and G. Molla. "Physiological functions of D-amino acid oxidases: From yeast to humans." In: *Cellular and Molecular Life Sciences* 64.11 (2007), pp. 1373–1394. DOI: 10.1007/s00018-007-6558-4.

[240] Rungaroon Waditee, Takashi Hibino, Yoshito Tanaka, Tatsunosuke Nakamura, Aran Incharoensakdi, and Teruhiro Takabe. "Halotolerant Cyanobacterium *Aphanothece halophytica* contains an Na^+/H^+ antiporter, homologous to eukaryotic ones, with novel ion specificity affected by C-terminal tail." In: *The Journal of Biological Chemistry* 276.40 (2001), pp. 36931–36938. DOI: 10.1074/jbc.M103650200.

[241] Renate Dippel and Winfried Boos. "The maltodextrin system of *Escherichia coli*: Metabolism and transport." In: *Journal of Bacteriology* 187.24 (2005), pp. 8322–8331. DOI: 10.1128/JB.187.24.8322-8331.2005.

[242] Sungmin Hwang, Kyoung Hwa Choi, Jieun Kim, and Jaeho Cha. "Biochemical characterization of 4-α-glucanotransferase from *Saccharophagus degradans* 2-40 and its potential role in glycogen degradation." In: *FEMS Microbiology Letters* 344.2 (2013), pp. 145–151. DOI: 10.1111/1574-6968.12167.

[243] H. M. Lam and M. E. Winkler. "Characterization of the complex pdxH-tyrS operon of *Escherichia coli* K-12 and pleiotropic phenotypes caused by pdxH insertion mutations." In: *Journal of Bacteriology* 174.19 (1992), pp. 6033–6045. DOI: 10.1128/jb.174.19.6033-6045.1992.

[244] Kristian Kvint, Laurence Nachin, Alfredo Diez, and Thomas Nyström. "The bacterial universal stress protein: Function and regulation." In: *Current Opinion in Microbiology* 6.2 (2003), pp. 140–145. DOI: 10.1016/S1369-5274(03)00025-0.

[245] Vivek Anantharaman, Lakshminarayan M. Iyer, and L. Aravind. "Ter-dependent stress response systems: Novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing." In: *Molecular BioSystems* 8.12 (2012), pp. 3142–3165. DOI: 10.1039/c2mb25239b.

[246] Alexis H.K. Choi, Leyla Slamti, Fikri Y. Avci, Gerald B. Pier, and Tomás Maira-Litrán. "The pgaABCD locus of *Acinetobacter baumannii* encodes the production of poly-β-1-6-N-acetylglucosamine, which is critical for biofilm formation." In: *Journal of Bacteriology* 191.19 (2009), pp. 5953–5963. DOI: 10.1128/JB.00647-09.

[247] Jason W Reed and Graham C Walker. "The exoD Gene of *Rhizobium meliloti* encodes a novel function needed for Alfalfa nodule invasion." In: *Journal of Bacteriology* 173.2 (1991), pp. 664–677. DOI: 10.1128/jb.173.2.664-677.1991.

[248] Yasser Y. Lenis, Mohammed A. Elmetwally, Juan G. Maldonado-Estrada, and Fuller W. Bazer. "Physiological importance of polyamines." In: *Zygote* 25.3 (2017), pp. 244–255. DOI: 10.1017/S0967199417000120.

[249] A. J. Michael. "Biosynthesis of polyamines and polyamine-containing molecules." In: *Biochemical Journal* 473.15 (2016), pp. 2315–2329. DOI: 10.1042/BCJ20160185.

[250] Seiya Watanabe, Masaki Yamada, Iwao Ohtsu, and Keisuke Makino. "α-Ketoglutaric semialdehyde dehydrogenase isozymes involved in metabolic pathways of D-glucarate, D-galactarate, and hydroxy-L-proline: Molecular and metabolic convergent evolution." In: *Journal of Biological Chemistry* 282.9 (2007), pp. 6685–6695. DOI: 10.1074/jbc.M611057200.

[251] Michal Koblížek. "Ecology of aerobic anoxygenic phototrophs in aquatic environments." In: *FEMS Microbiology Reviews* 39.6 (2015), pp. 854–870. DOI: 10.1093/femsre/fuv032.

[252] Bork Ansgar Berghoff, Jens Glaeser, Aaron Mischa Nuss, Monica Zobawa, Friedrich Lottspeich, and Gabriele Klug. "Anoxygenic photosynthesis and photooxidative stress: A particular challenge for Roseobacter." In: *Environmental Microbiology* 13.3 (2011), pp. 775–791. DOI: 10.1111/j.1462-2920.2010.02381.x.

[253] Christopher T. Jurgenson, Tadhg P. Begley, and Steven E. Ealick. "The structural and biochemical foundations of thiamin biosynthesis." In: *Annual Review of Biochemistry* 78.1 (2009), pp. 569–603. DOI: 10.1146/annurev.biochem.78.072407.102340.

[254] Raymond Wightman and Peter A. Meacock. "The THI5 gene family of *Saccharomyces cerevisiae*: Distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine." In: *Microbiology* 149.6 (2003), pp. 1447–1460. DOI: 10.1099/mic.0.26194-0.

[255] Amy Haas Jenkins, Ghislain Schyns, Sébastien Potot, Guangxing Sun, and Tadhg P. Begley. "A new thiamin salvage pathway." In: *Nature Chemical Biology* 3.8 (2007), pp. 492–497. DOI: 10.1038/nchembio.2007.13.

[256] Sergio A. Sa nudo-Wilhelmy, Laura Gómez-Consarnau, Christopher Suffridge, and Eric A. Webb. "The role of B-vitamins in marine biogeochemistry." In: *Annual Review of Marine Science* 6.1 (2014), pp. 339–367. DOI: 10.1146/annurev-marine-120710-100912.

[257] Irene Wagner-Döbler et al. "The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: A hitchhiker's guide to life in the sea." In: *ISME Journal* 4.1 (2010), pp. 61–77. DOI: 10.1038/ismej.2009.94.

[258] Semen A. Leyn, Inna A. Suvorova, Tatiana D. Kholina, Sofia S. Sherstneva, Pavel S. Novichkov, Mikhail S. Gelfand, and Dmitry A. Rodionov. "Comparative genomics of transcriptional regulation of methionine metabolism in proteobacteria." In: *PLoS ONE* 9.11 (2014). DOI: 10.1371/journal.pone.0113714.

[259] Julio C Gonzáles, Katrina Peariso, James E Penner-hahn, and Rowena G Matthews. "Cobalamin-independent methionine synthase from *Escherichia coli*: A Zinc Metalloenzyme." In: *Biochemistry* 35.38 (1996), pp. 12228–12234.

[260] Agnieszka Sekowska, Valérie Dénervaud, Hiroki Ashida, Karine Michoud, Dieter Haas, Akiho Yokota, and Antoine Danchin. "Bacterial variations on the methionine salvage pathway." In: *BMC Microbiology* 4 (2004), pp. 1–17. DOI: 10.1186/1471-2180-4-9.

[261] Eva Albers. "Metabolic characteristics and importance of the universal methionine salvage pathway recycling methionine from 5'-methylthioadenosine." In: *IUBMB Life* 61.12 (2009), pp. 1132–1142. DOI: 10.1002/iub.278.

[262] Jeffrey A Hughes. "In vivo hydrolysis of S-adenosyl-L-methionine in *Escherichia coli* increases export of 5-methylthioribose." In: *Canadian journal of microbiology* 52.6 (2006), pp. 599–602. DOI: 10.1139/w06-008.

[263] R. A. Bender. "Regulation of the histidine utilization (Hut) system in bacteria." In: *Microbiology and Molecular Biology Reviews* 76.3 (2012), pp. 565–584. DOI: 10.1128/MMBR.00014-12.

[264] Peep Palumaa. "Copper chaperones. the concept of conformational control in the metabolism of copper." In: *FEBS Letters* 587.13 (2013), pp. 1902–1910. DOI: 10.1016/j.febslet.2013.05.019.

[265] Andre Zapun, James C A Bardwell, and Thomas E Creighton J. "The reactive and destabilizing disulfide bond of DsbA , a protein required for protein disulfide bond formation in vivo." In: *Biochemistry* 32.19 (1993), pp. 5083–5092. DOI: 10.1021/bi00070a016.

[266] Melanie A. Adams and Zongchao Jia. "Structural and biochemical evidence for an enzymatic quinone redox cycle in *Escherichia coli*: Identification of a novel quinol monooxygenase." In: *Journal of Biological Chemistry* 280.9 (2005), pp. 8358–8363. DOI: 10.1074/jbc.M412637200.

[267] F. De Rienzo, R.R. Gabdoulline, M. C. Menziani, and R.C. Wade. "Blue copper proteins: A comparative analysis of their molecular interaction properties." In: *Protein Science* 9.8 (2000), pp. 1439–1454. DOI: 10.1110/ps.9.8.1439.

[268] Octavian Călinescu, Cristina Paulino, Werner Kühlbrandt, and Klaus Fendler. "Keeping it simple – Transport mechanism and pH regulation in Na^+/H^+ exchangers." In: *J Biol Chem* 289.19 (2014), pp. 13168–76. DOI: 10.1074/jbc.M113.542993.

[269] B. Aral and P. Kamoun. "The proline biosynthesis in living organisms." In: *Amino Acids* 13.3-4 (1997), pp. 189–217. DOI: 10.1007/BF01372588.

[270] E A Galinski. "Osmoadaptation in bacteria." In: *Advances in Microbial Physiology* 37 (1995), pp. 273–328. DOI: 10.1016/S0065-2911(08)60148-4.

[271] Ukti S Amin, Timothy D Lash, and Brian J Wilkinson. "Proline betaine is a highly effective osmoprotectant for *Staphylococcus aureus*." In: *Archives of Microbiology* 163.2 (1995), pp. 138–142. DOI: 10.1007/BF00381788.

[272] Christopher A. Dunn, Suzanne F. O'Handley, David N. Frick, and Maurice J. Bessman. "Studies on the ADP-ribose pyrophosphatase subfamily of the nudix hydrolases and tentative identification of trgB, a gene associated with tellurite resistance." In: *Journal of Biological Chemistry* 274.45 (1999), pp. 32318–32324. DOI: 10.1074/jbc.274.45.32318.

[273] Suzanne F. O'Handley, David N. Frick, Christopher A. Dunn, and Maurice J. Bessman. "Orf186 represents a new member of the nudix hydrolases, active on adenosine(5')triphospho(5')adenosine, ADP-ribose, and NADH." In: *Journal of Biological Chemistry* 273.6 (1998), pp. 3192–3197. DOI: 10.1074/jbc.273.6.3192.

[274] Elaine L. Jacobson, Daniel Cervantes-Laurean, and Myron K. Jacobson. "Glycation of proteins by ADP-ribose." In: *Molecular and Cellular Biochemistry* 138.1-2 (1994), pp. 207–212. DOI: 10.1007/BF00928463.

[275] Corinna Rademacher, Marie Christine Hoffmann, Jan Wilm Lackmann, Roman Moser, Yvonne Pfänder, Silke Leimkühler, Franz Narberhaus, and Bernd Masepohl. "Tellurite resistance gene *trgB* confers copper tolerance to *Rhodobacter capsulatus*." In: *BioMetals* 25.5 (2012), pp. 995–1008. DOI: 10.1007/s10534-012-9566-2.

[276] Ivan Erill, Susana Campoy, and Jordi Barbé. "Aeons of distress: An evolutionary perspective on the bacterial SOS response." In: *FEMS Microbiology Reviews* 31.6 (2007), pp. 637–656. DOI: 10.1111/j.1574-6976.2007.00082.x.

[277] Evelyn Barth, Katherine V. Gora, Katharina M. Gebendorfer, Florian Settele, Ursula Jakob, and Jeannette Winter. "Interplay of cellular cAMP levels, σS activity and oxidative stress resistance in *Escherichia coli*." In: *Microbiology* 155.5 (2009), pp. 1680–1689. DOI: 10.1099/mic.0.026021-0.

[278] Takuya Kasai, Atsushi Kouzuma, and Kazuya Watanabe. "CpdA is involved in amino acid metabolism in *Shewanella oneidensis* MR-1." In: *Bioscience, Biotechnology, and Biochemistry* (2017). DOI: 10.1080/09168451.2017.1413326.

[279] Iralis López-Villamizar, Alicia Cabezas, Rosa María Pinto, José Canales, Jo ao Meireles Ribeiro, José Carlos Cameselle, and María Jesús Costas. "The characterization of *Escherichia coli* CpdB as a recombinant protein reveals that, besides Having the expected 3-nucleotidase and 2,3-cyclic mononucleotide phosphodiesterase activities, it is also active as cyclic dinucleotide phosphodiesterase." In: *PLoS ONE* 11.6 (2016), pp. 1–22. DOI: 10.1371/journal.pone.0157308.

[280] Emilykate Mcdonough, Heather Kamp, and Andrew Camilli. "*Vibrio cholerae* phosphatases required for the utilization of nucleotides and extracellular DNA as phosphate sources." In: *Molecular Microbiology* 99.3 (2016), pp. 453–469. DOI: 10.1111/mmi.13128.

[281] Konrad Trülzsch, Andreas Roggenkamp, Cosima Pelludat, Alexander Rakin, Christoph A Jacobi, and Jürgen Heesemann. "Cloning and characterization of the gene encoding periplasmic 2´,3´-cyclic phosphodiesterase of *Yersinia enterocolitica* O:8." In: *Microbiology* 147 (2001), pp. 203–213. DOI: 10.1099/00221287-147-1-203.

[282] Hans Jörg Kunte. "Osmoregulation in bacteria: Compatible solute accumulation and osmosensing." In: *Environmental Chemistry* 3.2 (2006), pp. 94–99. DOI: 10.1071/EN06016.

[283] Anne Carlijn Alderkamp, Marion Van Rijssel, and Henk Bolhuis. "Characterization of marine bacteria and the activity of their enzyme systems involved in degradation of the algal storage glucan laminarin." In: *FEMS Microbiology Ecology* 59.1 (2007), pp. 108–117. DOI: 10.1111/j.1574-6941.2006.00219.x.

[284] Roberta Leonardi, Yong Mei Zhang, Charles O. Rock, and Suzanne Jackowski. "Coenzyme A: Back in action." In: *Progress in Lipid Research* 44 (2005), pp. 125–153. DOI: 10.1016/j.plipres.2005.04.001.

[285] Eugene V. Koonin. "Orthologs, paralogs, and evolutionary genomics." In: *Annual Review of Genetics* 39.1 (2005), pp. 309–338. DOI: 10.1146/annurev.genet.39.073003.114725.

[286] Leonidas Salichos and Antonis Rokas. "Inferring ancient divergences requires genes with strong phylogenetic signals." In: *Nature* 497.7449 (2013), pp. 327–331. DOI: 10.1038/nature12130.

[287] Dana E. Hunt, Lawrence A. David, Dirk Gevers, Sarah E. Preheim, Eric J. Alm, and Martin F. Polz. "Resource partitioning and sympatric differentiation among closely related bacterioplankton." In: *Science* 320.5879 (2008), pp. 1681–1685. DOI: 10.1126/science.1157890.

[288] Duccio Medini, Claudio Donati, Hervé Tettelin, Vega Masignani, and Rino Rappuoli. "The microbial pan-genome." In: *Current Opinion in Genetics and Development* 15.6 (2005), pp. 589–594. DOI: 10.1016/j.gde.2005.09.006.

[289] Todd J. Treangen and Eduardo P C Rocha. "Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes." In: *PLoS Genetics* 7.1 (2011). DOI: 10.1371/journal.pgen.1001284.

[290] H. Tettelin et al. "Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome"." In: *Proceedings of the National Academy of Sciences* 102.39 (2005), pp. 13950–13955. DOI: 10.1073/pnas.0506758102.

[291] Hervé Tettelin, David Riley, Ciro Cattuto, and Duccio Medini. "Comparative genomics: the bacterial pan-genome." In: *Current Opinion in Microbiology* 11.5 (2008), pp. 472–477. DOI: 10.1016/j.mib.2008.09.006.

[292] Mark Eppinger et al. "Genome sequences of the biotechnologically important *Bacillus megaterium* Strains QM B1551 and DSM319." In: *Journal of Bacteriology* 193.16 (2011), pp. 4199–4213. DOI: 10.1128/JB.00449-11.

[293] Tristan Lefébure and Michael J. Stanhope. "Evolution of the core and pan-genome of *Streptococcus*: Positive selection, recombination, and genome composition." In: *Genome Biology* 8.5 (2007), pp. 1–17. DOI: 10.1186/gb-2007-8-5-r71.

[294] Francesca Bottacini, Duccio Medini, Angelo Pavesi, Francesca Turroni, Elena Foroni, David Riley, Vanessa Giubellini, Hervé Tettelin, Douwe Van Sinderen, and Marco Ventura. "Comparative genomics of the genus *Bifidobacterium*." In: *Microbiology* 156.11 (2010), pp. 3243–3254. DOI: 10.1099/mic.0.039545-0.

[295] Adam C. Martiny, Kathleen Treseder, and Gordon Pusch. "Phylogenetic conservatism of functional traits in microorganisms." In: *ISME Journal* 7.4 (2013), pp. 830–838. DOI: 10.1038/ismej.2012.160.

[296] Konstantinos T Konstantinidis and James M Tiedje. "Trends between gene content and genome size in prokaryotic species with larger genomes." In: *Proceedings of the National Academy of Sciences of the United States of America* 9 (2004), pp. 3160–3165. DOI: 10.1073/pnas.0308653100.

[297] Antony T. Vincent, Mélanie V. Trudel, Luca Freschi, Vandan Nagar, Cynthia Gagné-Thivierge, Roger C. Levesque, and Steve J. Charette. "Increasing genomic diversity and evidence of constrained lifestyle evolution due to insertion sequences in *Aeromonas salmonicida*." In: *BMC Genomics* 17.44 (2016). DOI: 10.1186/s12864-016-2381-3.

[298] Federico M Lauro et al. "The genomic basis of trophic strategy in marine bacteria." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.37 (2009), pp. 15527–33. DOI: 10.1073/pnas.0903507106.

[299] Yuri I. Wolf and Eugene V. Koonin. "Genome reduction as the dominant mode of evolution." In: *BioEssays* 35.9 (2013), pp. 829–837. DOI: 10.1002/bies.201300037.

[300] Ariane Toussaint and Mick Chandler. "Prokaryote genome fluidity: Toward a system approach of the mobilome." In: *Bacterial Molecular Networks*. Ed. by Jacques van Helden, Ariane Toussaint, and Denis Thieffry. New York: Humana Press, 2012. Chap. 4, pp. 57–80. ISBN: 978-1-61779-360-8.

[301] Dan I. Andersson and Diarmaid Hughes. "Gene amplification and adaptive evolution in bacteria." In: *Annual Review of Genetics* 43.1 (2009), pp. 167–195. DOI: 10.1146/annurev-genet-102108-134805.

[302] Pavel Drevinek, Adam Baldwin, Laurens Lindenburg, Lovleen Tina Joshi, Angela Marchbank, Sarka Vosahlikova, Christopher G. Dowson, and Eshwar Mahenthiralingam. "Oxidative stress of *Burkholderia cenocepacia* induces insertion sequence-mediated genomic rearrangements that interfere with macrorestriction-based genotyping." In: *Journal of Clinical Microbiology* 48.1 (2010), pp. 34–40. DOI: 10.1128/JCM.01433-09.

[303] Colin J. Coros, Carol Lyn Piazza, Venkata R. Chalamcharla, Dorie Smith, and Marlene Belfort. "Global regulators orchestrate group II intron retromobility." In: *Molecular Cell* 34.2 (2009), pp. 250–256. DOI: 10.1016/j.molcel.2009.03.014.

[304] Cécile Pasternak, Bao Ton-Hoang, Geneviève Coste, Adriana Bailone, Michael Chandler, and Suzanne Sommer. "Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence." In: *PLoS Genetics* 6.1 (2010), pp. 1–10. DOI: 10.1371/journal.pgen.1000799.

[305] Elena Casacuberta and Josefa González. "The impact of transposable elements in environmental adaptation." In: *Molecular Ecology* 22.6 (2013), pp. 1503–1517. DOI: 10.1111/mec.12170.

[306] Daniel M. Stoebel and Charles J. Dorman. "The effect of mobile element IS10 on experimental regulatory evolution in *Escherichia coli*." In: *Molecular Biology and Evolution* 27.9 (2010), pp. 2105–2112. DOI: 10.1093/molbev/msq101.

[307] Alexandre M. Anesio and Christopher M. Bellas. "Are low temperature habitats hot spots of microbial evolution driven by viruses?" In: *Trends in Microbiology* 19.2 (2011), pp. 52–57. DOI: 10.1016/j.tim.2010.11.002.

[308] Ildefonso Cases, Victor De Lorenzo, and Christos A Ouzounis. "Transcription regulation and environmental adaptation in bacteria." In: *Trends in Microbiology* 11.6 (2003), pp. 248–253. DOI: 10.1016/S0966-842X(03)00103-3.

[309] Pablo A. Nu nez, Héctor Romero, Marisa D. Farber, and Eduardo P C Rocha. "Natural selection for operons depends on genome size." In: *Genome Biology and Evolution* 5.11 (2013), pp. 2242–2254. DOI: 10.1093/gbe/evt174.

[310] Llyd E. Wells and Jody W. Deming. "Modelled and measured dynamics of viruses in Arctic winter sea-ice brines." In: *Environmental Microbiology* 8.6 (2006), pp. 1115–1121. DOI: 10.1111/j.1462-2920.2006.00984.x.

[311] Jeff S. Bowman. "The relationship between sea ice bacterial community structure and biogeochemistry: A synthesis of current knowledge and known unknowns." In: 3 (2015). DOI: 10.12952/journal.elementa.000072.

[312] Martin T. Croft, Andrew D. Lawrence, Evelyne Raux-Deery, Martin J. Warren, and Alison G. Smith. "Algae acquire vitamin B12 through a symbiotic relationship with bacteria." In: *Nature* 438.7064 (2005), pp. 90–93. DOI: 10.1038/nature04056.

[313] Benjamin R.K. Roller, Steven F. Stoddard, and Thomas M. Schmidt. "Exploiting rRNA operon copy number to investigate bacterial reproductive strategies." In: *Nature Microbiology* 1.September (2016). DOI: 10.1038/nmicrobiol.2016.160.

[314] A. Boetius et al. "Export of algal biomass from the melting Arctic sea ice." In: *Science* 339.6126 (2013), pp. 1430–1432. DOI: 10.1126/science.1231346.

[315] Dorte Haubjerg Søgaard, Morten Kristensen, Søren Rysgaard, Ronnie Nøhr Glud, Per Juel Hansen, and Karen Marie Hilligsøe. "Autotrophic and heterotrophic activity in Arctic first-year sea ice: Seasonal study from Malene Bight, SW Greenland." In: *Marine Ecology Progress Series* 419 (2010), pp. 31–45. DOI: 10.3354/meps08845.

[316] Richard J. Geider and Julie La Roche. "Redfield revisited: Variability of C:N:P in marine microalgae and its biochemical basis." In: *European Journal of Phycology* 37.1 (2002), pp. 1–17. DOI: 10.1017/S0967026201003456.

[317] Alasdair M. Cook and Karin Denger. "Dissimilation of the C2 sulfonates." In: *Archives of Microbiology* 179.1 (2002), pp. 1–6. DOI: 10.1007/s00203-002-0497-0.

[318] Karin Denger, Jürgen Ruff, David Schleheck, and Alasdair M. Cook. "*Rhodococcus opacus* expresses the *xsc* gene to utilize taurine as a carbon source or as a nitrogen source but not as a sulfur source." In: *Microbiology* 150.6 (2004), pp. 1859–1867. DOI: 10.1099/mic.0.27077-0.

[319] Robert C. Fahey. "Glutathione analogs in prokaryotes." In: *Biochimica et Biophysica Acta - General Subjects* 1830.5 (2013), pp. 3182–3198. DOI: 10.1016/j.bbagen.2012.10.006.

[320] Johannes Hartl, Patrick Kiefer, Fabian Meyer, and Julia A. Vorholt. "Longevity of major coenzymes allows minimal de novo synthesis in microorganisms." In: *Nature Microbiology* 2.May (2017). DOI: 10.1038/nmicrobiol.2017.73.

[321] Qiang Zheng, Michal Koblížek, Thomas J Beatty, and Nianzhi Jiao. "Evolutionary divergence of marine aerobic anoxygenic phototrophic bacteria as seen from diverse organisations of their photosynthesis gene clusters." In: *Advances in Botanical Research Volume 66 – Genome Evolution of Photosynthetic Bacteria*. Ed. by Jean-pierre Jacquot, Pierre Gadal, and J. Thomas Beatty. Amsterdam: Elsevier Ltd., 2013. Chap. 12, pp. 359–383. ISBN: 9780123979230. DOI: 10.1016/B978-0-12-397923-0.00012-6.

[322] James N. Sturgis and Robert A. Niederman. "Organization and assembly of light-harvesting complexes in the purple bacterial membrane." In: *Advances in Photosynthesis and Respiration - Volume 28: The Purple Phototrophic Bacteria*. Ed. by Neil C. Hunter, Fevzi Daldal, Marion C. Thurnauer, and Thomas J. Beatty. 1. Springer, 2009. Chap. 14, pp. 253–273. ISBN: 978-1-4020-8814-8.

[323] Merja Suutari and Siino Laakso. "Microbial fatty acids and thermal adaptation." In: *Critical Reviews in Microbiology* 20.4 (1994), pp. 285–328. DOI: 10.3109/10408419409113560.

[324] Michael Hügler and Stefan M. Sievert. "Beyond the calvin cycle: Autotrophic carbon fixation in the ocean." In: *Annual Review of Marine Science* 3.1 (2011), pp. 261–289. DOI: 10.1146/annurev-marine-120709-142712.

[325] Thibaud Taillefumier, Anna Posfai, Yigal Meir, and Ned S Wingreen. "Microbial consortia at steady supply." In: *eLife* 6 (2017). Ed. by Carl T Bergstrom, e22644. DOI: 10.7554/eLife.22644.

[326] George N. Somero. "Adaptation of enzymes to temperature: Searching for basic "strategies"." In: *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology* 139 (2004), pp. 321–333. DOI: 10.1016/j.cbpc.2004.05.003.

[327] Raghu Prasad Rao Metpally and Boojala Vijay B. Reddy. "Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins." In: *BMC Genomics* 10.11 (2009). DOI: 10.1186/1471-2164-10-11.

[328] Alexandra Moura, Michael A. Savageau, and Rui Alves. "Relative amino acid composition signatures of organisms and environments." In: *PLoS ONE* 8.10 (2013), e77319. DOI: 10.1371/journal.pone.0077319.

[329] Emmanuele De Vendittis, Immacolata Castellano, Roberta Cotugno, Maria Rosaria Ruocco, Gennaro Raimo, and Mariorosario Masullo. "Adaptation of model proteins from cold to hot environments involves continuous and small adjustments of average parameters related to amino acid composition." In: *Journal of Theoretical Biology* 250.1 (2008), pp. 156–171. DOI: 10.1016/j.jtbi.2007.09.006.

[330] David K. Smith, Predrag Radivojac, Zoran Obradovic, A. Keith Dunker, and Guang Zhu. "Improved amino acid flexibility parameters." In: *Protein Science* 12.5 (2003), pp. 1060–1072. DOI: 10.1110/ps.0236203.

[331] Nikolaj Tindbaek, Allan Svendsen, Peter Rahbek Oestergaard, and Henriette Draborg. "Engineering a substrate-specific cold-adapted subtilisin." In: *Protein Engineering, Design and Selection* 17.2 (2004), pp. 149–156. DOI: 10.1093/protein/gzh019.

[332] Shinsuke Maki, Mizuho Yoneta, and Yasuhiro Takada. "Two isocitrate dehydrogenases from a psychrophilic bacterium, *Colwellia psychrerythraea*." In: *Extremophiles* 10.3 (2006), pp. 237–249. DOI: 10.1007/s00792-005-0493-9.

[333] Hye Yeon Koh, Hyun Park, Jun Hyuck Lee, Se Jong Han, Young Chang Sohn, and Sung Gu Lee. "Proteomic and transcriptomic investigations on cold-responsive properties of the psychrophilic Antarctic bacterium *Psychrobacter* sp. PAMC 21119 at subzero temperatures." In: *Environmental Microbiology* 19.2 (2017), pp. 628–644. DOI: 10.1111/1462-2920.13578.

[334] Federico M. Lauro, Khiem Tran, Alessandro Vezzi, Nicola Vitulo, Giorgio Valle, and Douglas H. Bartlett. "Large-scale transposon mutagenesis of *Photobacterium profundum* SS9 reveals new genetic loci important for growth at low temperature and high pressure." In: *Journal of Bacteriology* 190.5 (2008), pp. 1699–1709. DOI: 10.1128/JB.01176-07.

[335] James J. Foti, Lyle A. Simmons, Penny J. Beuning, and Graham C. Walker. "Signal transduction in the *Escherichia coli* SOS response." In: *Handbook of Cell Signaling*. Second Edition. Elsevier Inc., 2010. Chap. 258, pp. 2127–2136. ISBN: 9780123741455. DOI: 10.1016/B978-0-12-374145-5.00258-8.

[336] L. Marsh and G. C. Walker. "Cold sensitivity induced by overproduction of UmuDC in *Escherichia coli*." In: *Journal of Bacteriology* 162.1 (1985), pp. 155–161.

[337] Carole L. Linster, Emile Van Schaftingen, and Andrew D. Hanson. "Metabolite damage and its repair or pre-emption." In: *Nature Chemical Biology* 9.2 (2013), pp. 72–80. DOI: 10.1038/nchembio.1141.

[338] Tatiana V. Karpinets et al. "Conserved synteny at the protein family level reveals genes underlying *Shewanella* species' cold tolerance and predicts their novel phenotypes." In: *Functional and Integrative Genomics* 10.1 (2010), pp. 97–110. DOI: 10.1007/s10142-009-0142-y.

[339] K. Limsuwun and P. G. Jones. "Spermidine acetyltransferase is required to prevent spermidine toxicity at low temperatures in *Escherichia coli*." In: *Journal of Bacteriology* 182.19 (2000), pp. 5373–5380. DOI: 10.1128/JB.182.19.5373-5380.2000.

[340] Jon M Kaguni. "Replication initiation at the *Escherichia coli* chromosomal origin." In: *Current Opinion in Chemical Biology* 15.5 (2011), pp. 606–613. DOI: 10.1016/j.cbpa.2011.07.016.

[341] Ricardo Cavicchioli. "Microbial ecology of Antarctic aquatic systems." In: *Nature Reviews Microbiology* 13.11 (2015), pp. 691–706. DOI: 10.1038/nrmicro3549.

[342] B. J. Tindall. "Prokaryotic diversity in the Antarctic: The tip of the iceberg." In: *Microbial Ecology* 47.3 (2004), pp. 271–283. DOI: 10.1007/s00248-003-1050-7.

[343] Liisa B. Koski and G. Brian Golding. "The closest BLAST hit is often not the nearest neighbor." In: *Journal of Molecular Evolution* 52.6 (2001), pp. 540–542. DOI: 10.1007/s002390010184.

[344] Seungtai Yoon, Zhenyu Xuan, Vladimir Makarov, Kenny Ye, and Jonathan Sebat. "Sensitive and accurate detection of copy number variants using read depth of coverage." In: *Genome Research* 19.9 (2009), pp. 1586–1592. DOI: 10.1101/gr.092981.109.

DECLARATION

I intend to write a doctoral thesis on the topic Distribution and ecological characteristics of members of the *Roseobacter* group at Georg-August-Universität Göttingen. I will be supervised by Prof. Dr. Rolf Daniel.

I hereby declare that:

1. the opportunity to work on the aforementioned doctoral thesis project was not arranged commercially. Especially, I did not engage any organisation which searches for doctoral thesis supervisors or which will entirely or partly carry out my examination duties against payment;

2. I have only accepted and will only accept the assistance of third parties in so far as it is scientifically justifiable and acceptable in regards to the examination regulations. Especially, all parts of the dissertation will be written by myself; I have not accepted and will not accept impermissible help from other parties neither for money nor for free;

3. I will observe the Statute of the Georg-August-University Göttingen for ensuring good scientific practice;

4. I have not applied for corresponding doctoral degree procedures at any other university in Germany or abroad; the submitted doctoral thesis or parts thereof were not used in another doctoral degree procedure.

I am aware that incorrect information precludes the admission to doctoral studies resp. may later on lead to the discontinuation of the doctoral degree procedures or to the revocation of the doctoral degree.

Göttingen, 24/06/2020

Florian Lenk

CURRICULUM VITAE

Florian Lenk
* 01.11.1986

✉ Quantzstraße 32a
37127 Scheden

✉ flenk@gwdg.de

since 02/2015	doctorate in the Dept. of Genomic and Applied Microbiology, Faculty of Biology and Psychology, Georg-August University Göttingen
10/2009 – 09/2014	M.Sc. in Microbiology and Biochemistry, Georg-August University Göttingen
10/2006 – 09/2009	B.Sc. in Molecular Medicine, Georg-August University Göttingen
07/1999 – 06/2006	university-entrance diploma at the Grotefend Gymnasium Hann. Münden

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Final Version as of June 24, 2020 (`classicthesis` version 4.2).