

Discovery and Functional Characterization of Novel Soil-metagenome Derived Phosphatases

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr.rer.nat.)

of the Georg-August-Universität Göttingen

within the doctoral program Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Genis Andrés Castillo Villamizar

from Bucaramanga, Colombia

Göttingen, 2019

Thesis Committee

Prof. Dr. Rolf Daniel, Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

PD Dr. Michael Hoppert, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Members of the Examination Board

Referent: **Prof. Dr. Rolf Daniel**, Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Coreferent: **PD Dr. Michael Hoppert**, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Further Members of the Examination Board

Prof. Dr. Stefanie Pöggeler, Department of Genetics of Eukaryotic Microorganisms, Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Fabian Commichau, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Kai Heimel Department of Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Gerhard Braus, Department of Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Wolfgang Streit, Department of Microbiology and Biotechnology, Institute of Plant Science and Microbiology, University of Hamburg

Date of oral examination: 28.03.2019

List of Publications

- 1. Function-Based Metagenomic Library Screening and Heterologous Expression Strategy for Genes Encoding Phosphatase Activity**
Genis A. Castillo Villamizar, Heiko Nacke, Rolf Daniel. *Methods Mol Biol.* (2017). 1539:249-260. DOI: 10.1007/978-1-4939-6691-2_16
- 2. Functional Metagenomics Reveals an Overlooked Diversity and Novel Features of Soil-Derived Bacterial Phosphatases and Phytases**
Genis Andrés Castillo Villamizar, Heiko Nacke, Marc Boehning, Kristin Herz, Rolf Daniel. *mbio* (2019). DOI: 10:e01966-18, DOI:10.1128/mBio.01966-18
- 3. Characteristics of the First Protein Tyrosine Phosphatase with Phytase Activity from a Soil Metagenome**
Genis Andrés Castillo Villamizar, Heiko Nacke, Laura Griese, Lydia Tabernero, Katrina Funkner and Rolf Daniel. *Genes* (2019). 10:101. DOI: 10.3390/genes10020101
- 4. Functional metagenomics unwraps a new catalytic domain associated to phytase activity: the metallo- β -lactamase superfamily domain**
Genis Andrés Castillo Villamizar, Katrina Funkner, Heiko Nacke, Karolin Foerster and Rolf Daniel. accepted *mSphere* DOI: 10.1128/mSphere.00167-19
- 5. Fine Spatial Scale Variation of Soil Microbial Communities under European Beech and Norway Spruce**
Heiko Nacke, Kezia Goldmann, Ingo Schöning, Birgit Pfeiffer, Kristin Kaiser, Genis Andrés Castillo-Villamizar, Marion Schrupf, François Buscot, Rolf Daniel, Tesfaye Wubet. Fine spatial scale variation of soil microbial communities under european beech and norway spruce. *Front Microbiol* (2017):2067. DOI: 10.3389/fmicb.2016.02067
- 6. Genome Sequence of the Acetogenic Bacterium *Moorella mulderi* DSM 14980T**
Genis Andrés Castillo-Villamizar, Anja Poehlein. *Genome Announc.* (2016). 4, e00444-00416. DOI: 10.1128/genomeA.00444-16
- 7. First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic *Sporomusa sphaeroides* Strain E (DSM 2875)**
Genis Andrés Castillo-Villamizar, Rolf Daniel, Anja Poehlein. *Genome Announc.* (2017). 4, e00444-00416. DOI: 10.1128/genomeA.00037-17

Table of Contents

Thesis Committee	I
List of Publications	II
1 General Introduction	1
1.1. Phosphorus “the essential element”	1
1.2. Factors Disturbing the Phosphorus Cycle	1
1.3. Phytate as Alternative Source of Phosphorus	5
1.4. Phytases are Key Biocatalysts with Still Unknown Potential	6
1.5. Histidine Acid Phytases.....	9
1.6. β -propeller Phytases	11
1.7. Protein Tyrosin Phytases	12
1.8. Purple Acid Phytases.....	13
1.9. Soil Metagenomes as Source of Novel Phytases	14
1.10. Aim of the Thesis.....	16
1.11. References.....	18
2 Function-Based Metagenomic Library Screening and Heterologous Expression Strategy for Genes Encoding Phosphatase Activity	24
3 Functional Metagenomics Reveals an Overlooked Diversity and Novel Features of Soil-Derived Bacterial Phosphatases and Phytases	37
Supplemental Information for Chapter 3.....	53
4 Characteristics of the First Protein Tyrosine Phosphatase with Phytase Activity from a Soil Metagenome	60
Supplemental Information for Chapter 4.....	77

5	Functional Metagenomics Unwraps a New Catalytic Domain Associated to Phytase Activity: the Metallo-β-lactamase Superfamily Domain	81
5.1.	Abstract.....	82
5.2.	Importance	82
5.3.	Introduction	83
5.4.	Results	84
5.5.	Discussion	92
5.6.	References.....	101
5.7.	Supplemental Information for Chapter 5	105
6	General Discussion	117
6.1.	Soil as Source of Metagenome-Derived Biocatalysts.....	117
6.2.	Phosphatases in Metagenome Surveys	118
6.3.	Metagenomics for the Recovery of New Phytases.....	119
6.4.	A Simple and Effective Function-Based Method for the Retrieval of Novel Phosphatases/Phytases.....	123
6.5.	The Metagenome Input to Phytase Research	126
6.6.	Function-Based Metagenomics Reveals Classic Phytases with Special Characteristics and New Phosphatases	127
6.7.	References.....	130
7	Summary.....	134
8	Appendix.....	136
8.1.	Thesis Declaration.....	136
8.2.	Additional Publications.....	137
8.2.1.	Fine Spatial Scale Variation of Soil Microbial Communities under European Beech and Norway Spruce.....	137
	Supplemental Information for Chapter 8.2.1	152

8.2.2. Genome Sequence of the Acetogenic Bacterium <i>Moorella mulderi</i> DSM 14980T.....	167
8.2.3. First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic <i>Sporomusa sphaeroides</i> Strain E (DSM 2875).....	166
8.3. Acknowledgments.....	169
8.4. Curriculum Vitae	171

1 General Introduction

1.1. *Phosphorus “the essential element”*

“We may be able to substitute nuclear power for coal, and plastics for wood, and yeast for meat, and friendliness for isolation—but for phosphorus, there is neither substitute nor replacement” (Isaac Asimov 1974) (1).

Life has different levels of dependency on some key elements such as carbon (C) nitrogen (N) or phosphorus (P). The dependence level varies with the element in terms of quality, quantity or bioavailability. P is particularly interesting because it is considered the element that most strongly limits the biological productivity of the entire planet (2). Compared with other key chemical elements P is a rare resource in the biosphere. With respect to mass terms, P does not rank among the first 10 most abundant elements neither on land nor in water. The P mass value on earth ranks between that of aluminum and chloride. Furthermore, it has been estimated that the earth’s crust contains 1,200 mg P kg⁻¹. In plants, associated P concentration varies from approximately 0.1 mg P L⁻¹ in soil solution to near 4,000 mg P kg⁻¹ in seeds. Remarkably, despite its scarcity in the biosphere, P is relatively abundant in vertebrate bodies. Mammals contain around 25 g P kg⁻¹ making phosphorus the second most abundant essential mineral after calcium in this type of organisms (3, 4).

P is present in numerous substances essential for biological and biogeochemical processes. Nevertheless, P involvement in life processes does not occur in its elemental form. P is highly reactive and quickly converts to phosphate or other oxidized forms. P plays a role in numerous biological processes, including energy production, metabolism and bone mineralization. Additionally, it provides the structural framework for DNA, RNA and phospholipids. (3, 5, 6). All living organisms require regular P intake and so far synthetic substitutes are not known. Thus, without phosphorus, no life is possible (2). The high level of biological relevance of P, has created an increasing interest in its cycling, origin, sources and processing.

1.2. *Factors Disturbing the Phosphorus Cycle*

Different biotic factors play a role in P cycling. For instance, microorganisms decompose dead biomass and solubilize otherwise unavailable soil phosphates such as some types of

inositol phosphates (7). Unlike other cycles such as C and N cycle, the P cycle is not driven by biotic factors. Physical transfers are limited, as P does not form long-lived gaseous compounds. Therefore, the amount of P that can be stored in the atmosphere is rather low. Without relevant atmospheric connection, P is then mainly restricted to solid and liquid phases. The moderate solubility, makes P mobility very limited compared to other elements such as N or C (2, 3).

Phosphate rocks represent the majority of the usable P on earth's crust. Most of the other forms in which P exists are unavailable for direct biological uptake. P transfer into terrestrial systems from phosphate rocks is performed through processes such as weathering or leaching. These processes are slow and a major constraint with respect to terrestrial primary productivity. (8, 9).

By using the civilization time (10^3 years) as timescale, the general natural global P cycle seems slow and unidirectional. First, phosphate rocks are transferred into the ocean in form of soluble and particulate P. Next, the transferred P sinks into the sediments where it is slowly recycled and re-exposed as phosphate rocks by the reshaping of the Earth's surface. All of that occurs in about 10^7 to 10^8 years. The cycle shows only minor interruptions, mostly due to temporary absorption of a reduced fraction of the transferred P by biotic factors. However, the fraction of the cycle corresponding to organic P cycling (land- and water-based) has significantly shorter times scales (10^{-2} to 10^0 years). Other land-based cycles in smaller scales transport phosphates from soils to plants with a subsequent return of a large portion of P to soils due to the mineralization of plant litter, dead microorganisms and other biomass sources (Fig.1) (2) .

Substantial input of P from atmospheric deposition does not occur in P cycle compared with other key elements, e.g. N. The loss of P in the fast soil-plant sub-cycling is replaced exclusively by weathering of P-bearing rocks, which is a slow process (2, 3, 10).

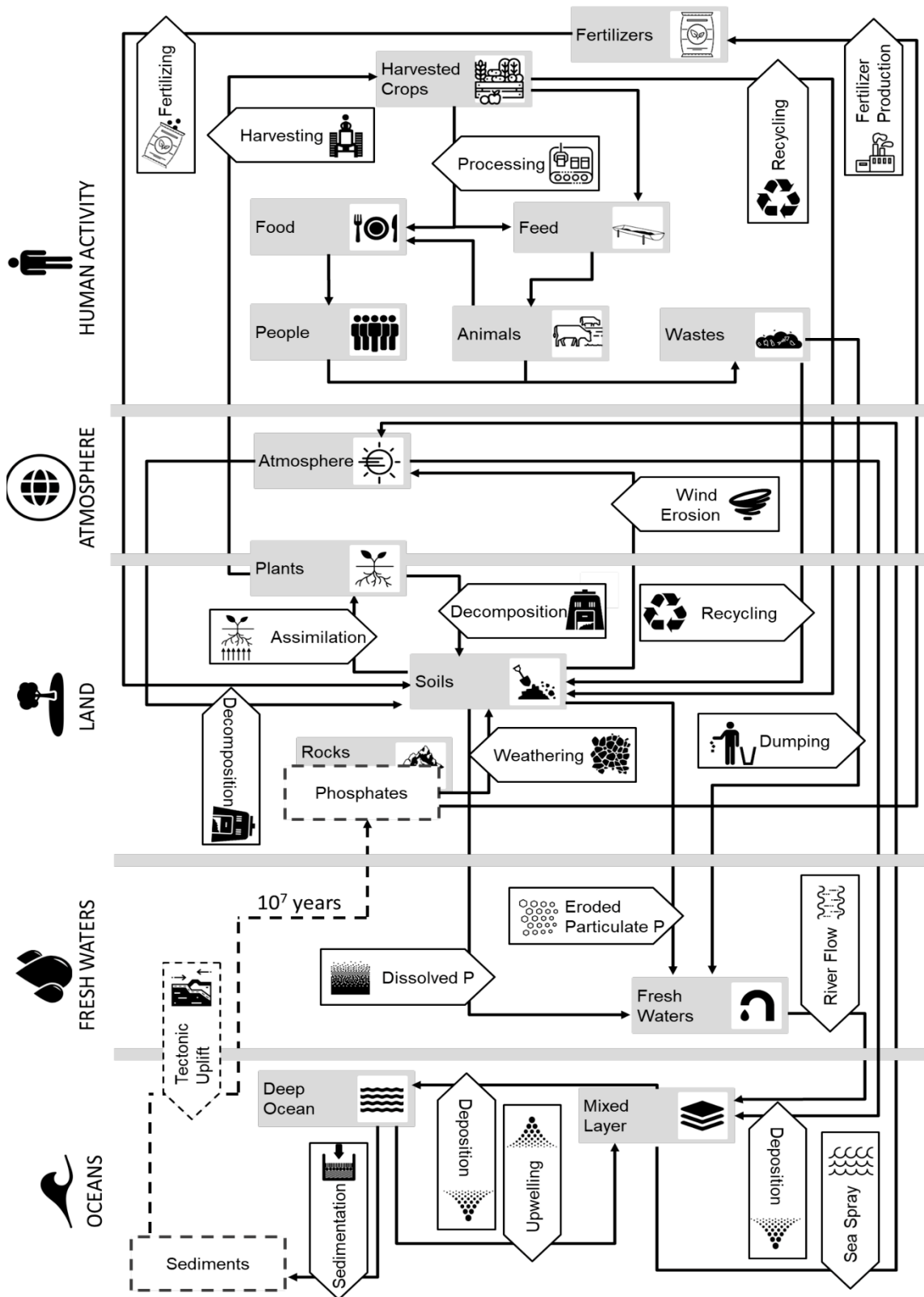


Fig. 1. Flux of phosphorus according to Smil 2000. Modified from (2)

The main disturbance arises from the anthropogenic impact on every major biogeochemical cycle (11). In the case of P, within the past century mankind has

quadrupled the environmental flow of phosphorus inputs to the biosphere. The main reason for this increase is the mining of phosphorus compounds for use as fertilizers (11). This has disturbed the P cycle and caused serious problem. We have too much and too little phosphorus at the same time (12). There is an excess of P contaminating the environment. This triggers a series of ecological problems, i.e. the eutrophication of freshwater bodies and oceans. At the same time, the shortage of mineable resources of P is a major challenge for the future of the global food demand (12-14).

Phosphate rock is a non-renewable natural resource and the minable deposits are limited to a few countries controlling 85% of the world's reserves. Thus, different groups around the world have warned about an upcoming P depletion.

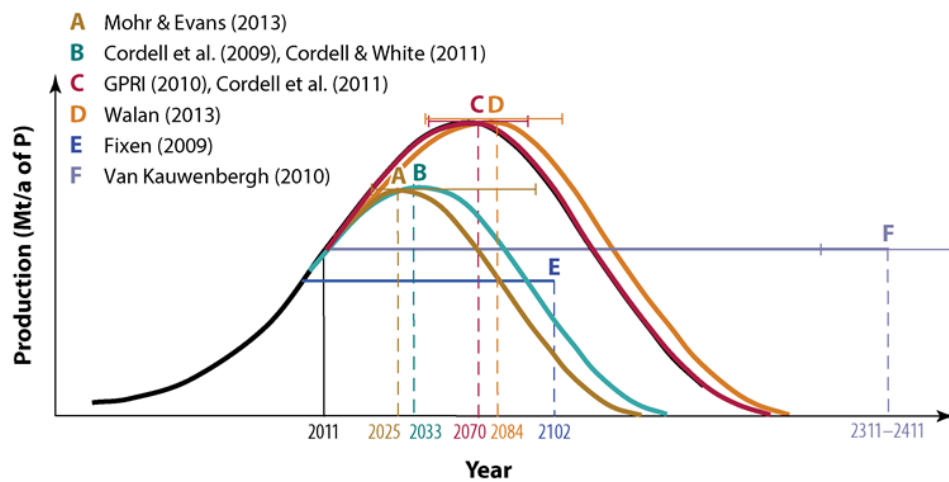


Fig. 2. Global phosphate depletion time scenarios by different studies. Modified from (10) .

Cordell and White, 2014 analyzed different published models of P production, usage and depletion. In their analysis the most optimistic models predict that in approximately 50 years the peak of the P production will be reached and the production reduction will begin (Fig. 2). Other less optimistic models consider that the peak of P production will be reached in the next 6-10 years. Despite the differences in the predicted times, it is accepted that P production is a highly relevant matter for mankind and measures need to be implemented in order to fulfill the future alimentary needs of the growing global population (10, 12, 15, 16).

The second part of the problem is related to the excess of P in some environments such as soils and water bodies. This excess of P originates also from human activities. Therefore, strategies directed to recapturing and recycling P as well as its redistribution from zones

with excessive concentration to zones with scarcity are being designed and implemented (12, 17).

In the early 2000s, more than 7 million tons/year of P were released into the environment. Most of this phosphorus derives from animal manure and excreta related to mass food production (18). An important portion of the excreted P is present in the form of unprocessed *myo*-inositol hexakisphosphate (InsP₆) also known as phytate in its salt form (Fig. 3) (14, 19). Phytate is considered an alternative source of P for the upcoming global needs.

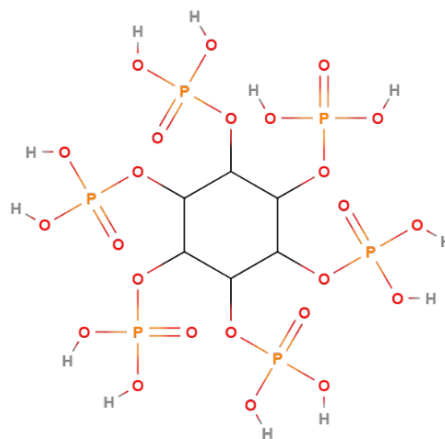


Fig. 3. 2D structure of phytic acid drawn with PubChem Sketcher (20).

1.3. Phytate as Alternative Source of Phosphorus

Myo-Inositol phosphates can contain between one and eight phosphoryl groups. This type of molecules is ubiquitous in eukaryotic species and has diverse biological functions. Phytate is the most abundant *myo*-Inositol phosphate and has multiple important roles in eukaryotic cellular processes such as DNA repair, RNA processing, mRNA export, plant development, apoptosis and pathogenicity (21-27).

Phytate is quantitatively the most important inositol phosphate found in soils and represents the main P storage molecule in cereals and grains (28, 29). Furthermore, phytate is considered a large and unexploited reservoir of phosphorus. More than 50 million tons of phytate are present yearly in commercial fruits and seeds (30-32). This amount is the equivalent to circa 70% of the annual worldwide phosphate (P) use with respect to mineral fertilizers (33). In animal diets containing oilseeds, legumes and cereals 60 to 80% of P is represented by phytate (29). Unfortunately, P in phytate form has a low bioavailability for monogastric species including several mass production

animals such as swine, poultry or fish (34). This low bioavailability makes phytate a contaminating P waste and results, i.e., in eutrophication of rivers and lakes. The processing of phytate is dependent on the presence of phytate degrading enzymes (phytases) that hydrolyze phytate. Phytases are considered valuable tools for the recovery and recycling of P. These enzymes improve cell ability to scavenge phosphorus from phytate present in different habitats including soils, water, manure, as well as excretions, and are industrially relevant biocatalysts (35-37).

1.4. Phytases are Key Biocatalysts with Still Unknown Potential

Life on earth is largely dependent on processes such as phosphorylation and dephosphorylation, which are actively regulating other processes including energy production, general metabolism and signal transduction (38).

The dephosphorylation process is achieved by phosphatases, which possess a high diversity and complexity. Different from kinases, phosphatases evolutionary history is not based on common ancestors. This led to a lower number of different phosphatase genes compared to that of kinase genes (38-40). Nevertheless, phosphatases hydrolyze a variety of substrates including phytate. Phytases degrade phytate to inositol phosphates and inorganic monophosphates. The term phytase is related specifically to the class of phosphatases with the *in vitro* capability to release at least one phosphate from phytate. Phytate-degrading enzymes have a wide distribution in plants, microorganisms and some animal tissues (41, 42).

In higher plants, e.g. legumes, oilseeds and nuts, phytases are present in grains, seeds and pollen. Phytase activity has been found in maize, tomato, wheat, barley, rye, spelt, canola, beans and *Arabidopsis* (43-45). Low levels of phytase activity have been reported in liver and blood of calves and the brush border vesicles of poultry. Interestingly, phytase activity can be also found in the mucosa extracts of the small intestine of rats, rabbits, guinea pigs, chicken, and humans. However, the level of activity of the enzymes in these tissues is extremely low and does not play a significant role in phytate-processing (46-49).

Microbial phytases have been detected in bacteria and fungi (33, 50). In bacteria, some of the best-studied phytases are derived from *Pseudomonas* sp., *Bacillus* sp., and *Klebsiella* sp. Bacteria isolates with phytase activity have been identified in very diverse

environments ranging from the extreme cold conditions of the Himalayan soils to hydrothermal aquatic environments (41, 51, 52).

Phytases from fungi are mostly associated to the histidine acid phosphatase family and exhibit high levels of glycosylated residues probably associated to protein stability. Moreover, differences in the catalytic mechanism between phytases from Ascomycetes and Basidiomycetes have been reported. The analysis of the phylogenetic relationships of fungal phytases have evidenced that phytases from Ascomycetes and their Basidiomycetes counterparts are separated into two different clades (41). Different species of *Aspergillus* such as *A. niger*, *A. fumigatus*, *A. terreus*, and *A. oryzae*, have been reported as belonging to the most efficient producers of extracellular phytases (53).

In the last three decades, phytases have been in the focus because of their potential for solving several of the previously mentioned problems. Phytases are considered as tools for reducing the impact of animal agriculture on the environment (36, 54). Phytases are used as animal feed supplements, enhancing the utilization of phytate-derived phosphorus and the phytate-bound minerals by monogastric animals, and reducing phosphorus excretion up to 50% (55). Probably one of the most important applications of phytases is related to their usage with respect to crops. Under the scenarios of P depletion and the subsequent limitations for crop production, phytases are powerful tools for P recovering. Expression of extracellular phytase from *Medicago truncatula* in transgenic *Arabidopsis* have revealed a significant improvement in organic phosphorus utilization and subsequent plant growth (56). Other potential applications for phytases include the *myo*-inositol phosphate synthesis for pharmaceutical purposes and their usage for human nutrition and disease prevention. The presence of phytate in food affects the bioavailability of many divalent and trivalent mineral ions such as Ca^{2+} , Mg^{2+} , Mn^{2+} , Zn^{2+} , $\text{Fe}^{2+/3+}$ and Cu^{2+} (37, 57). Therefore, phytases might help improving the bioavailability of these ions and thereby increasing the nutritional value of different foods. Yeasts with high phytase activity are already candidates for use in the manufacture of whole meal bread with high mineral bioavailability (58)

In addition to the direct application of phytase technologies, several phytases have been recently associated with processes of pathogenesis in different organisms. For instance, some phytases of plant pathogens manipulate signaling pathways in plants. Others stimulate intracellular head kidney leukocyte activities in fish. Additionally, phytases are

enhancers of the virulence potency of *Candida albicans*, and act as a counteracting force of intracellular growth restriction of *Legionella pneumophila* by phytic acid (59-61).

The study of and the search for novel and improved phytases resulted in identification of several types of phytases with distinct characteristics and phylogenetic relationships (62).

Depending on the position within the inositol ring, where the dephosphorylation is initiated, phytases can be grouped in 3 classes: The 3-phytases, 5-phytases and 6-phytases, which produce 1,2,4,5,6-pentakisphosphate, 1,2,3,4,6-pentakisphosphate and 1,2,3,4,5- pentakisphosphate, respectively (41). Phosphatases with phytase activity have different catalytic signatures and structures (Fig. 4). Based on the active site geometry, specific sequence patterns and catalytic mechanisms, phytases have been grouped into four classes: Histidine acid phytases (HAPhys), β -propeller phytases (BPPhys), protein tyrosine phytases (PTPhys), and purple acid phytases (PAPhys)(62, 63).

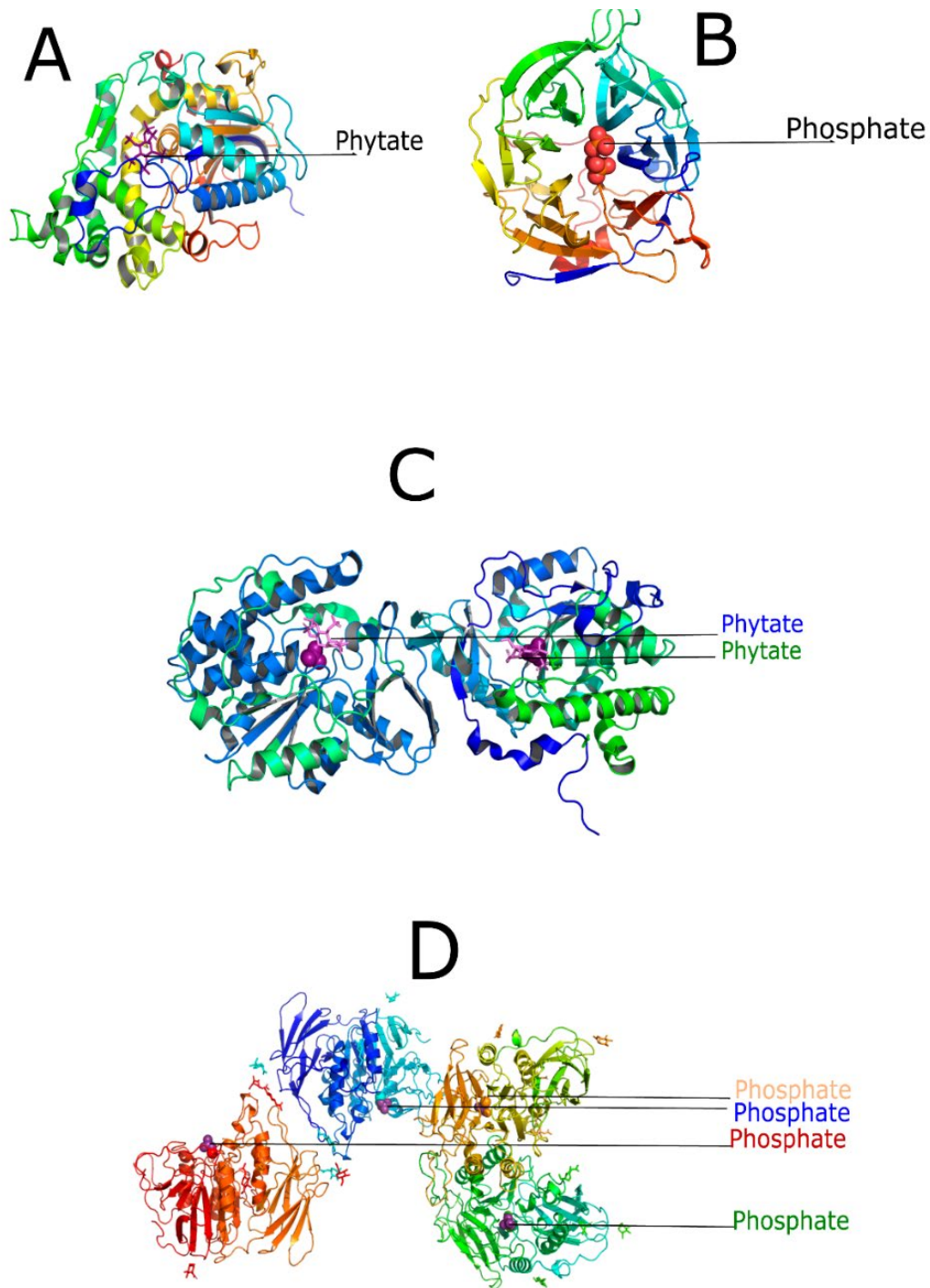


Fig. 4. Crystal structures (secondary structure) of representatives of each of the four structural classes of phosphatase with phytase activity. Images are not shown to scale. A) HAPhy, *Escherichia coli* (1DKQ) AppA in complex with phytate. B) BPPHy, *Bacillus amyloliquefaciens* (3AMR) phytase in complex with phosphate. C) PTPphy, *Selenomonas ruminantium* (3MMJ) in complex with phytate. D) PAPphy, *Phaseolus vulgaris* (4KBP) in complex with phosphate. Images created with Jalview 2.0 (64) using Protein Data Bank accessions as indicated.

1.5. Histidine Acid Phytases

HAPhy is a large class of phytases present in animals, plants and microorganisms. Fig. 5 depicts the general reaction of the HAPhys. Members of this class share a highly conserved

active site, an N terminal motif (RHGX_RXP) and a C-terminal motif (HD) (65). Both motifs are required to form a single catalytic center enabling a two-step mechanism for dephosphorylation of phytic acid. Firstly, the guanidinium group of arginine in the RHG peptide of the active site interacts with the phosphate group of the substrate, making the guanidinium group more susceptible for nucleophilic attack (66, 67). The aspartate residue of the C terminal motif is used as proton donor for the oxygen of the phosphomonoester bond. Although all enzymes of this class share this catalytic mechanism, the efficiency for phytate processing is highly variable among the enzymes (34).

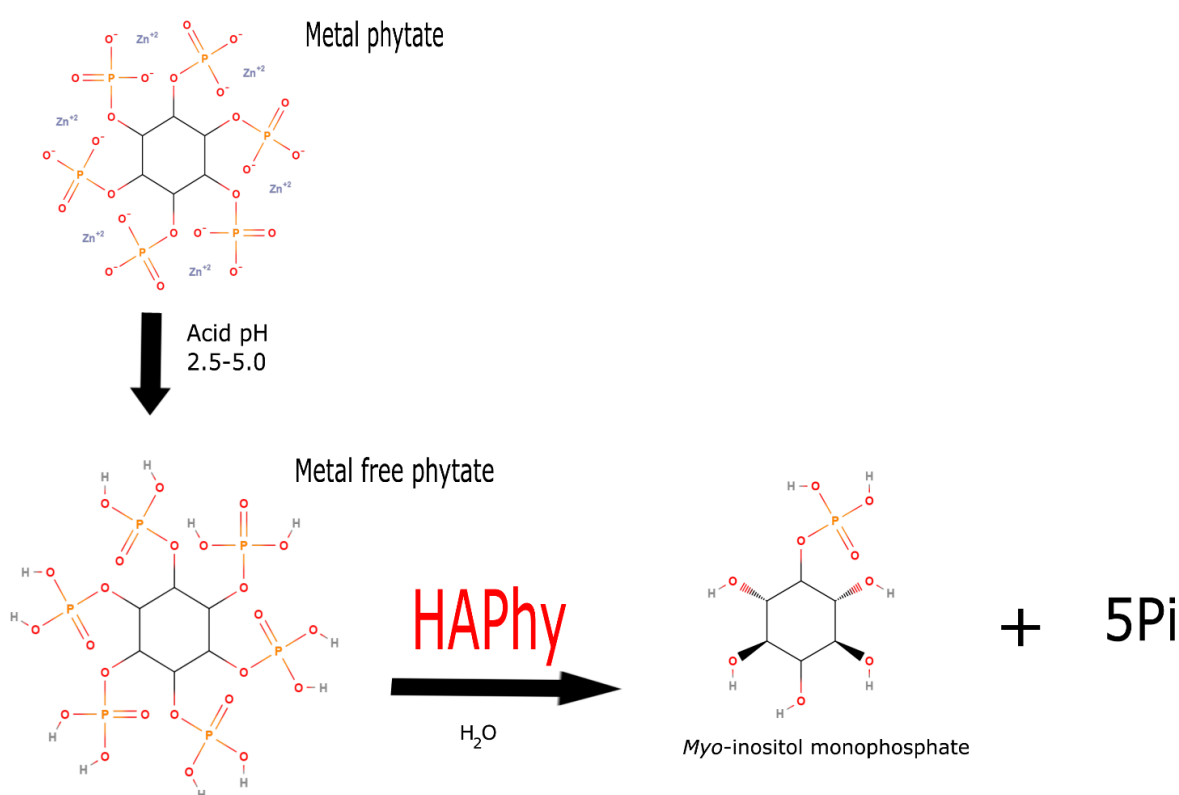


Fig. 5. Schematic illustration of substrate hydrolysis by histidine acid phytases. HAPhys can hydrolyze phytate at acidic pH, when phytate exists as metal-free phytate. Adapted from (67).

Since not all histidine acid phosphatases are phytases, it is clear that the ability to use phytate as substrate does not rely on a single catalytic feature. Only HAPhys can process phytate as substrate. Homology analysis between prokaryotic and eukaryotic HAPhys has shown low sequence similarity apart from their shared specific catalytic signatures. HAPhys can be divided into two groups, based on their substrate specificity (67): One group with narrow substrate specificity and high specific activity for phytate, and another group with a broad substrate specificity but low specific activity towards phytate.

Analysis of the catalytic sites of the HAPhys derived from *A. niger* and *A. fumigatus* indicated the importance of other amino acid residues with respect to the specificity towards phytate. In this sense position 300 was reported as highly variable, while position 301 was very conserved comprising a lysine. Variations in the amino acid residues present in position 300 of the HAPhy resulted in drastic changes of the specificity levels. In this regard, the change of a neutral amino acid for an acidic or basic one changed the specificity of the enzyme towards phytate from low to high (34, 68).

The HAPhy are the most extensively studied and commercialized phytases. Some applications of HAPhys include food-related applications such as production of phytate-free soybean milk, low-phytin bread. A different type of commercial utilization of HAPhys is connected to the production of genetically modified pigs also called Enviropig™. This type of animals possess salivary glands capable of produce phytase which improves phytate processing reducing phosphorus manure, and requirements of food additives (63).

1.6. β -propeller Phytases

BPPhys are also called alkaline phytases as their optimal activities are above pH 8.0. This type of phytases represent a class of enzymes that exhibits little or no significant homology or phylogenetic relationship to any other known phosphatase (69). The first representatives of BPPhy (PhyC and TS-Phy) were cloned from *Bacillus subtilis* and *B. amyloliquefaciens*, respectively (70, 71). The name was derived from their molecular structure, which is composed mainly of β -propeller sheets with a six bladed propeller (72) (Fig. 4B)

The catalytic model proposed by Shin et al. 2001 suggests the existence of cleavage and affinity sites. Two adjacent phosphate groups occupy both sites. Phosphate-binding to the affinity site enables the cleavage of the flanking phosphate by the cleavage site. The catalytic product of BPPhys is *myo*-inositol trisphosphate preferring hydrolysis of every second phosphate. One of the few similarities between BPPhys and other phosphatases was described by using a multidimensional index tree method for side-chain patterns. This analysis revealed that BPPhys and pyrophosphatases share some common structural features, including the cleavage and affinity sites (69). BPPhys are Ca^{2+} -dependent phytases. This cation Ca^{2+} serves as an essential activator of the enzyme and increases its thermostability (73).

BPPhy is the phytase family with most representatives in prokaryotes. BPPHys harbor specific domain structures and conserved cysteine residues. These features are used to classify BBPhys into several subgroups (I-VII). (74). BPPHys possess several characteristics of interest for biotechnological applications. Thermostability of the BPPHys is one desired characteristic for animal food processing. Other representatives improving body weight and phosphorus utilization efficiency of birds. (75). Moreover, the BPPhy from *Bacillus amyloliquefaciens* DS11 enhances the availability of minerals at neutral pH in phytate rich foods, like soymilk, wheat bread and others (76, 77). In addition to the biotechnological interest, some members of this phytase class have been related to pathogenic processes. *Xanthomonas oryzae* is a pathogen of rice and possesses the six-bladed β -propeller protein PhyA, which act as virulence factor (78).

1.7. Protein Tyrosin Phytases

PTPhy are protein tyrosine phosphatases (PTPs) with phytase activity. This type of enzymes was first reported during the late 1990s and is responsible for phytase activity of anaerobic ruminal bacteria (79). Chu et al. 2004 were the first who established the association of phytase activity with the classic PTPs (80). PTPhy possess the same catalytic mechanism and signature as PTPs. The characteristic PTP CX₅R(S/T) sequence pattern forms a loop for coordinating the scissile phosphate for nucleophilic attack (81-83). PTPhy activity is accomplished by two steps. A conserved cysteine residue required for activity acts as a thiolate and forms a phospho-cysteine intermediate (84). The conserved arginine residue stabilizes the scissile phosphate through the action of guanidinium group, which together with other nitrogens of the main chain stabilizes the highly negative charge of the phosphate group and coordinates the oxygen atoms into an optimal orientation (85). The catalytic function is dependent on the formation of a substrate-binding pocket. At the same time, the substrate specificity seems to be determined by the depth of the pocket (80, 86). The PTPhy model (Fig. 6) was established by analyzing the crystal structure of a phytase of *Selenomonas ruminantium* (80). The protein folds in two domains. One of the domains is larger than the other one and consists of four β -sheet strands flanked by α -helices on both sides. The smaller domain consists of a 5-stranded β -sheet. The phytate-binding pocket is formed at the interface of the two domains with the P loop at the bottom of the pocket. (80, 81)

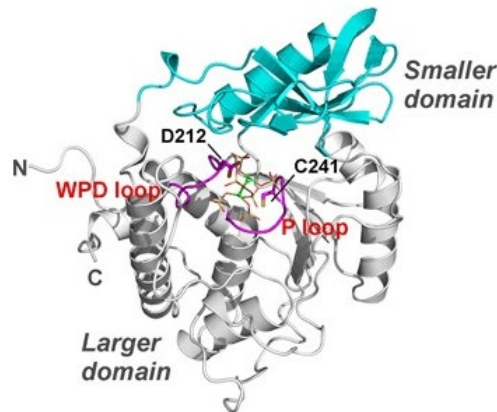


Fig. 6. Structure of PTPphy from *Selenomonas ruminantium* (PDB code: 1U26). The large and small domains of the enzyme are colored in gray and cyan, respectively. The side chains of the catalytic residues are presented as line and stick models. The loops that accommodate catalytic residues are indicated and colored in magenta. Modified from (87).

PTPs are involved in multiple regulatory functions and play a role in the cellular equilibrium of protein tyrosine phosphorylation by dephosphorylating tyrosine residues of proteins. PTPs also participate in cell signaling by dephosphorylating proteins at other amino acid residues (serine and threonine) or lipid substrates. Several members of this class have been associated with pathogenesis (88-91). In the specific case of the PTPphy the physiological function is not fully clear but similar to the classic PTPs some members of the PTPphy are also related to pathogenesis (33, 92).

1.8. Purple Acid Phytases

The purple acid phosphatases (PAPs) are part of the metallophosphoesterase superfamily. These proteins are present in animals, plants, bacteria and fungi (93). PAPs have diverse biological roles. For instance, in mammals the most important function of PAPs is related to increased bone resorption. Other functions involve iron transport, the generation of reactive oxygen species (ROS) and participation in the immune response (94, 95). In several plant species, PAPs expression is up-regulated during phosphate-scarcity, indicating that this type of enzymes helps the plant to overcome the P limitation (96-98). PAPs are recognized for hydrolyzing several phosphorylated compounds and some of the isoforms are also capable to hydrolyze phytic acid (PAPhy) (93). The first reported PAPhy (GmPhy) was identified in the cotyledons of a germinating soybean (99). More recently, other PAPhy have been identified in wheat, barley, maize and rice (45). In prokaryotes, homologous genes of PAPhy have been detected, but only one putative bacterial gene from metagenomic origin has been characterized and is described in the

chapter 3 of this study (100, 101). The structural analysis of the PAPhy is limited to extrapolation by comparing them with the structures of general PAPs.

Since low availability of P in soils hampers plant growth, the evolution of mechanisms to improve soil P availability is a key factor for many plant species. The genes *HvPAPhy_a*, *HvPAPhy_b1*, *HvPAPhy_b2*, *PHY_US417*, *PHYA*, *AVPIDOX*, *MtPHY1*, *MtPT1*, and *SK-57* are main genes evaluated in transgenic plants and microorganisms with reported phytase activity, and presumably are involved at least partially in P recovering for plants (102). Dionisio et. al. 2011 performed a phylogenetic analysis comprising 43 PAPs. All PAPhy members from wheat, barley, maize and rice grouped together with a collection of plant PAPs. The exception was the PAPhy from *Arabidopsis*, which grouped separately demonstrating the variability of the sequence (45). No PAPhy is currently commercialized, although genes *HvPAPhy* and *GmPAP14* have shown to increase phytase activity in barley and soybean, respectively. Improved phytate processing in different tissues of transgenic plants, indicates the potential of the PAPhy for plant production improvement (103, 104).

More than 110 years after the discovery of the first phytase (105), there are still many aspects associated with these enzymes that remain elusive to us. For instance, the *in vivo* function of most phytate degrading enzymes remain unknown. Only few PAPhy involved in seed germination are considered to have a real phytase activity proved *in vivo*, meanwhile for the vast majority of enzymes with phytate degrading capabilities the *in vivo* functionalities are highly speculative (106).

1.9. Soil Metagenomes as Source of Novel Phytases

An important aspect in the field of phytase research concerns the exploration of a wider range of phytate-degrading enzymes by using culture-independent techniques, e.g. screening of metagenomes or metagenomic libraries. Until now, with very few exceptions, the reported phytases and all commercial phytases were derived exclusively from cultured individual microorganisms or plants (33, 41, 106, 107).

From the industrial and environmental perspectives, the access to phytases associated to the non-cultivable fraction of microorganisms bears a high potential with respect to the search for novel phytases with improved characteristics (108). Similarly, from the ecological, evolutive and physiological perspectives, accessing the phytases of the non-culturable fraction increases our understanding of key aspects in the phosphatase field,

including the diversity, functionality, phylogenetic relationships or the pathogenic potential of environmentally-derived phosphatases/phytases (100, 109, 110).

Although prospection soil metagenomes resulted in retrieval and characterization of many types of biocatalysts such as cellulases (111), lipases (112), esterases (113), amylases (114) or proteases (115) among many others (116, 117), this is not the case for enzymes with phosphatase/phytase activity. Thus, the application of metagenome surveys, for the identification and characterization of functional phytases has become a cutting-edge topic in the phytase research field.

Although phosphatases/phytases are very important biocatalyst, function-based metagenomics has not produced as much success retrieving new types of enzymes from environmental samples as for other types of biocatalysts e.g. lipases or cellulases. In the particular case of the phytases, in the best of my knowledge before this study only two reports using functional metagenomics have been published. Nevertheless, only three genes were identified and one was expressed and its product characterized (110, 118).

1.10. Aim of the Thesis

Function-based metagenomics is a powerful tool, which has contributed to the study of gene function and to the discovery of novel biocatalysts with scientific and biotechnological relevance. Although numerous types of biocatalysts have been discovered and analyzed by function-based metagenomics, many others remain locked in environmental samples and attached to their non-culturable biological hosts. The latter is also the case for phosphatases and within this type for phytate-degrading enzymes. Phosphatases/phytases are key biocatalysts, which are involved in phosphorus processing at different levels. From cellular metabolism to soil decontamination phosphatases/phytases impact largely our way of life. Almost all existing information about phosphatases and particularly for phytases is based on enzymes derived from the culturable fraction of microorganisms. One of the main reasons for that is the scanty number of effective culture-independent methods for the retrieval of this type of biocatalysts. This signifies a critical knowledge gap in comparison with other important biocatalysts. The general aim of this study was to contribute to close that knowledge gap. In order to do so, three different concatenated goals were established and accomplished (Fig. 7).

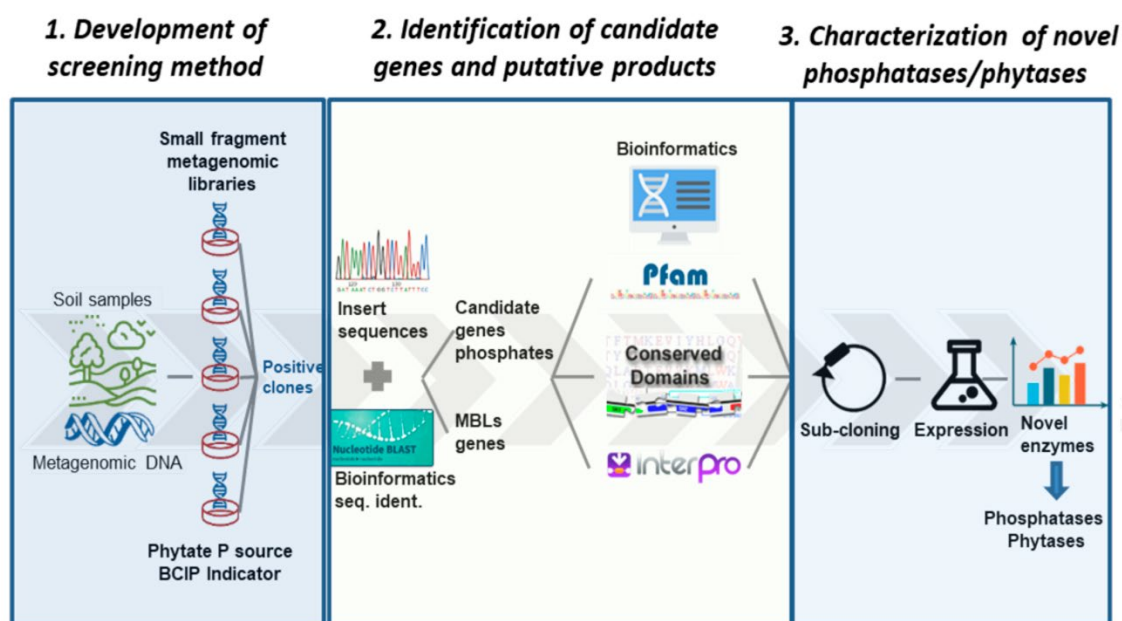


Fig. 7. General scheme of the study.

The first goal was the development and application of a simple and reliable function-based screening method for the identification of active clones with phosphatase/phytase

activity. The function-based screening was based on the evaluation of small-insert metagenomic libraries, derived from soil samples, and the use of phytate as source of phosphorous and inductor of the phytase activity.

After the effective recovery of multiple positive clones, the next objective was to identify putative phosphatase/phytase genes responsible for the phenotype of the recovered clones. This part of the strategy involved sequence identity comparisons against public databases and a subsequent search of specific phosphatase/phytase catalytic signatures in the products encoded by all candidate genes.

The last objective was the retrieval and characterization of selected gene products, in order to determinate their functionality *in vitro*. This strategy included sub-cloning of candidate genes, followed by the biochemical profiling of the products. The selection of the candidates for protein characterization was made following criteria such as novelty of the sequences and the presence/absence of specific catalytic signatures in the encoded products.

1.11. References

1. Asimov I. 1975. Life's Bottleneck, p 170, Asimov on chemistry Anchor Books, Doubleday, Garden city, New York.
2. Smil V. 2000. Phosphorus in the environment: Natural flows and human interferences. *Annu Rev Energ Env* 25:53-88.
3. Tiessen H. 2008. Phosphorus in the global environment, p 1-7. *In* White PJ, Hammond JP (ed), *The ecophysiology of plant-phosphorus interactions* doi:10.1007/978-1-4020-8435-5_1. Springer Netherlands, Dordrecht.
4. McDowell LR. 2003. Chapter 2 - calcium and phosphorus, p 33-100. *In* McDowell LR (ed), *Minerals in Animal and Human Nutrition (Second Edition)* doi:10.1016/B978-0-444-51367-0.50005-2. Elsevier, Amsterdam.
5. Rauchfuss H, Mitchell TN. 2008. "Chemical Evolution", p 87-124. *In* Rauchfuss H, Mitchell TN (ed), *Chemical Evolution and the Origin of Life* doi:10.1007/978-3-540-78823-2_4. Springer Berlin Heidelberg, Berlin, Heidelberg.
6. Westheimer FH. 1987. Why nature chose phosphates. *Science* 235:1173-1178.
7. Sharma SB, Sayyed RZ, Trivedi MH, Gobi TA. 2013. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. *Springerplus* 2:587.
8. Suh S, Yee S. 2011. Phosphorus use-efficiency of agriculture and food system in the US. *Chemosphere* 84:806-813.
9. Raven JA. 2008. Phosphorus and the future, p 271-283. *In* White PJ, Hammond JP (ed), *The Ecophysiology of Plant-Phosphorus Interactions* doi:10.1007/978-1-4020-8435-5_12. Springer Netherlands, Dordrecht.
10. Cordell D, White S. 2014. Life's bottleneck: Sustaining the world's phosphorus for a food secure future. *Annu Rev Env Resour* 39:161-188.
11. Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Höglberg P, Linder S, Mackenzie FT, Moore III B, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W. 2000. The global carbon cycle: A test of our knowledge of earth as a system. *Science* 290:291-296.
12. Elser J, Bennett E. 2011. Phosphorus cycle: A broken biogeochemical cycle. *Nature* 478:29-31.
13. Childers DL, Corman J, Edwards M, Elser JJ. 2011. Sustainability challenges of phosphorus and food: Solutions from closing the human phosphorus cycle. *BioScience* 61:117-124.
14. Scholz RW, Ulrich AE, Eilittä M, Roy A. 2013. Sustainable use of phosphorus: a finite resource. *Sci Total Environ* 461-462:799-803.
15. Cordell D, Rosemarin A, Schröder JJ, Smit AL. 2011. Towards global phosphorus security: A systems framework for phosphorus recovery and reuse options. *Chemosphere* 84:747-758.
16. Geissler B, Mew MC, Weber O, Steiner G. 2015. Efficiency performance of the world's leading corporations in phosphate rock mining. *Resour Conserv Recy* 105:246-258.
17. Schröder JJ, Smit AL, Cordell D, Rosemarin A. 2011. Improved phosphorus use efficiency in agriculture: a key requirement for its sustainable use. *Chemosphere* 84:822-831.
18. Cordell D, Drangert J-O, White S. 2009. The story of phosphorus: Global food security and food for thought. *Global Environ Chang* 19:292-305.
19. Giles CD, Cade-Menun BJ. 2014. Phytate in animal manure and soils: Abundance, cycling and bioavailability, p 163-190. *In* He Z, Zhang H (ed), *Applied Manure and Nutrient Chemistry for Sustainable Agriculture and Environment* doi:10.1007/978-94-017-8807-6_9. Springer Netherlands, Dordrecht.
20. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE. 2019. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 47:D1102-D1109.
21. Sasakawa N, Sharif M, Hanley MR. 1995. Metabolism and biological activities of inositol pentakisphosphate and inositol hexakisphosphate. *Biochem Pharmacol* 50:137-146.

22. Irvine RF, Schell MJ. 2001. Back in the water: the return of the inositol phosphates. *Nat Rev Mol Cell Biol* 2:327.
23. York JD, Odom AR, Murphy R, Ives EB, Wentz SR. 1999. A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 285:96-100.
24. Hanakahi LA, Bartlett-Jones M, Chappell C, Pappin D, West SC. 2000. Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. *Cell* 102:721-729.
25. Macbeth MR, Schubert HL, VanDemark AP, Lingam AT, Hill CP, Bass BL. 2005. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309:1534-1539.
26. Majerus PW, Zou J, Marjanovic J, Kisseleva MV, Wilson MP. 2008. The role of inositol signaling in the control of apoptosis. *Adv Enzyme Regul* 48:10-17.
27. Gillaspay GE. 2011. The cellular language of *myo*-inositol signaling. *New Phytol* 192:823-839.
28. Turner BL, Cheesman AW, Godage HY, Riley AM, Potter BVL. 2012. Determination of neo- and D-chiro-inositol hexakisphosphate in soils by solution ³¹P NMR spectroscopy. *Environ Sci Technol* 46:4994-5002.
29. Ravindran V, Ravindran G, Sivalogan S. 1994. Total and phytate phosphorus contents of various foods and feedstuffs of plant origin. *Food Chem* 50:133-136.
30. Gerke J. 2015. Phytate (inositol hexakisphosphate) in soil and phosphate acquisition from inositol phosphates by higher plants. A review. *Plants* 4:253-266.
31. Lott JNA, Ockenden I, Raboy V, Batten GD. 2007. Phytic acid and phosphorus in crop seeds and fruits: A global estimate. *Seed Sci Res* 10:11-33.
32. Lei XG, Porres JM. 2007. Phytase and inositol phosphates in animal nutrition: Dietary manipulation and phosphorus excretion by animals, p 133-149. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment* doi:10.1017/S001447970700587X. CABI, Wallingford, UK.
33. Mullaney E, Ullah AH. 2007. Phytases: Attributes, catalytic mechanisms and applications, p 97-110. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.
34. Mullaney EJ, Daly CB, Ullah AHJ. 2000. Advances in phytase research, p 157-199, *Adv Appl Microbiol*, vol Volume 47. Academic Press.
35. Vats P, Bhattacharyya MS, Banerjee UC. 2005. Use of phytases (*myo*-inositolhexakisphosphate phosphohydrolases) for combatting environmental pollution: a biological approach. *Crit Rev Env Sci Tec* 35:469-486.
36. Golovan SP, Meidinger RG, Ajakaiye A, Cottrill M, Wiederkehr MZ, Barney DJ, Plante C, Pollard JW, Fan MZ, Hayes MA, Laursen J, Hjorth JP, Hacker RR, Phillips JP, Forsberg CW. 2001. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat Biotechnol* 19:741-745.
37. Kumar A, Chanderman A, Makolomakwa M, Perumal K, Singh S. 2016. Microbial production of phytases for combating environmental phosphate pollution and other diverse applications. *Crit Rev Env Sci Tec* 46:556-591.
38. Kennelly PJ. 2001. Protein phosphatases - A phylogenetic perspective. *Chem Rev* 101:2291-2312.
39. Fahs S, Lujan P, Köhn M. 2016. Approaches to study phosphatases. *ACS Chem Biol* 11:2944-2961.
40. Chen MJ, Dixon JE, Manning G. 2017. Genomics and evolution of protein phosphatases. *Sci Signal* 10:eaag1796.
41. Ushasree MV, Vidya J, Pandey A. 2017. 14 - other enzymes: phytases, p 309-333. *In* Pandey A, Negi S, Soccol CR (ed), *Current Developments in Biotechnology and Bioengineering*. Elsevier, Amsterdam, Netherlands.
42. Mullaney EJ, Ullah AHJ. 2003. The term phytase comprises several different classes of enzymes. *Biochem Bioph Res Co* 312:179-184.

43. Bouain N, Rouached A, Prom-u-thai C, Hanin M, Pandey AK, Rouached H. 2017. Phosphate, phytate and phytases in plants: from fundamental knowledge gained in *Arabidopsis* to potential biotechnological applications in wheat. *Crit Rev Biotechnol* 37:898-910.
44. Li M, Osaki M, Honma M, Tadano T. 1997. Purification and characterization of phytase induced in tomato roots under phosphorus-deficient conditions. *Soil Sci Plant Nutr* 43:179-190.
45. Dionisio G, Madsen CK, Holm PB, Welinder KG, Jorgensen M, Stoger E, Arcalis E, Brinch-Pedersen H. 2011. Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice. *Plant Physiol* 156:1087-1100.
46. McCollum EV, Hart EB. 1908. On the occurrence of a phytin-splitting enzyme in animal tissues. *J Biol Chem* 4:497-500.
47. Bitar K, Reinhold JG. 1972. Phytase and alkaline phosphatase activities in intestinal mucosae of rat, chicken, calf, and man. *Biochim Biophys Acta* 268:442-452.
48. Iqbal TH, Lewis KO, Cooper BT. 1994. Phytase activity in the human and rat small intestine. *Gut* 35:1233-1236.
49. Ellestad LE, Angel R, Soares JH. 2002. Intestinal phytase II: A comparison of activity and *in vivo* phytate hydrolysis in three teleost species with differing digestive strategies. *Fish Physiol Biochem* 26:259-273.
50. Konietzny U, Greiner R. 2002. Molecular and catalytic properties of phytate-degrading enzymes (phytases). *Int J Food Sci Tech* 37:791-812.
51. Kalsi HK, Singh R, Dhaliwal HS, Kumar V. 2016. Phytases from *Enterobacter* and *Serratia* species with desirable characteristics for food and feed applications. *3 Biotech* 6:64.
52. Jorquera MA, Gabler S, Inostroza NG, Acuña JJ, Campos MA, Menezes-Blackburn D, Greiner R. 2018. Screening and characterization of phytases from bacteria isolated from Chilean hydrothermal environments. *Microb Ecol* 75:387-399.
53. Shieh TR, Ware JH. 1968. Survey of microorganism for the production of extracellular phytase. *Appl Microbiol* 16:1348-1351.
54. Zhang X, Li Z, Yang H, Liu D, Cai G, Li G, Mo J, Wang D, Zhong C, Wang H, Sun Y, Shi J, Zheng E, Meng F, Zhang M, He X, Zhou R, Zhang J, Huang M, Zhang R, Li N, Fan M, Yang J, Wu Z. 2018. Novel transgenic pigs with enhanced growth and reduced environmental impact. *Elife* 7:e34286.
55. Lei XG, Ku PK, Miller ER, Yokoyama MT. 1993. Supplementing corn-soybean meal diets with microbial phytase linearly improves phytate phosphorus utilization by weanling pigs. *J Anim Sci* 71:3359-3367.
56. Xiao K, Harrison MJ, Wang Z-Y. 2005. Transgenic expression of a novel *M. truncatula* phytase gene results in improved acquisition of organic phosphorus by *Arabidopsis*. *Planta* 222:27-36.
57. Kumar V, Sinha AK, Makkar HPS, Becker K. 2010. Dietary roles of phytate and phytase in human nutrition: A review. *Food Chem* 120:945-959.
58. Nuobariene L, Hansen ÅS, Arneborg N. 2012. Isolation and identification of phytase-active yeasts from sourdoughs. *Lwt-Food Sci Technol* 48:190-196.
59. Weber S, Stirnimann CU, Wieser M, Frey D, Meier R, Engelhardt S, Li X, Capitani G, Kammerer RA, Hilbi H. 2014. A type IV translocated *Legionella* cysteine phytase counteracts intracellular growth restriction by phytate. *J Biol Chem* 289:34175-34188.
60. Lazado CC, Caipang CMA, Gallage S, Brinchmann MF, Kiron V. 2010. Responses of Atlantic cod *Gadus morhua* head kidney leukocytes to phytase produced by gastrointestinal-derived bacteria. *Fish Physiol Biochem* 36:883-891.
61. Tsang PW, Fong WP, Samaranyake LP. 2017. *Candida albicans* orf19.3727 encodes phytase activity and is essential for human tissue damage. *PLoS One* 12:e0189219.
62. Lei XG, Weaver JD, Mullaney E, Ullah AH, Azain MJ. 2013. Phytase, a new life for an "old" enzyme. *Annu Rev Anim Biosci* 1:283-309.
63. Lei XG, Porres JM, Mullaney EJ, Brinch-Pedersen H. 2007. Phytase: Source, structure and application, p 505-529. *In* Polaina J, MacCabe AP (ed), *Industrial Enzymes: Structure, Function and Applications* doi:10.1007/1-4020-5377-0_29. Springer Netherlands, Dordrecht.

64. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189-1191.
65. Ostanin K, Harms EH, Stevis PE, Kuciel R, Zhou MM, Van Etten RL. 1992. Overexpression, site-directed mutagenesis, and mechanism of *Escherichia coli* acid phosphatase. *J Biol Chem* 267:22830-22836.
66. Mullaney EJ, Ullah AHJ. 1998. Conservation of the active site motif in *Aspergillus niger* (*ficuum*) pH 6.0 optimum acid phosphatase and kidney bean purple acid phosphatase. *Biochem Biophys Res Commun* 243:471-473.
67. Oh BC, Choi WC, Park S, Kim YO, Oh TK. 2004. Biochemical properties and substrate specificities of alkaline and histidine acid phytases. *Appl Microbiol Biotechnol* 63:362-372.
68. Wyss M, Brugger R, Kronenberger A, Rémy R, Fimbel R, Oesterhelt G, Lehmann M, van Loon APGM. 1999. Biochemical characterization of fungal phytases (*myo*-inositol hexakisphosphate phosphohydrolases): catalytic properties. *Appl Environ Microbiol* 65:367-373.
69. Shin S, Ha N-C, Oh B-C, Oh T-K, Oh B-H. 2001. Enzyme mechanism and catalytic property of β propeller phytase. *Structure* 9:851-858.
70. Kim Y-O, Kim H-K, Bae K-S, Yu J-H, Oh T-K. 1998. Purification and properties of a thermostable phytase from *Bacillus* sp. DS11. *Enzyme Microb Tech* 22:2-7.
71. Cheng C, Lim BL. 2006. Beta-propeller phytases in the aquatic environment. *Arch Microbiol* 185:1-13.
72. Ha NC, Kim YO, Oh TK, Oh BH. 1999. Preliminary x-ray crystallographic analysis of a novel phytase from a *Bacillus amyloliquefaciens* strain. *Acta Crystallogr D Biol Crystallogr* D55.
73. Kerovuo J, Lappalainen I, Reinikainen T. 2000. The metal dependence of *Bacillus subtilis* phytase. *Biochem Biophys Res Commun* 268:365-369.
74. Lim BL, Yeung P, Cheng C, Hill JE. 2007. Distribution and diversity of phytate-mineralizing bacteria. *ISME J* 1:321-330.
75. Rao DECS, Rao KV, Reddy VD. 2008. Cloning and expression of *Bacillus phytase* gene (phy) in *Escherichia coli* and recovery of active enzyme from the inclusion bodies. *J Appl Microbiol* 105:1128-1137.
76. Kwon D-A, Park S, Kwon D, Kim K-H, Oh B-C, Auh J-H. 2014. Improving mineral availability in soymilk by dephosphorylation of phytic acid using an alkaline phytase from *Bacillus amyloliquefaciens* DS11. *Food Sci Biotechnol* 23:1067-1072.
77. Park Y-J, Park J, Park K-H, Oh B-C, Auh J-H. 2011. Supplementation of alkaline phytase (Ds11) in whole-wheat bread reduces phytate content and improves mineral solubility. *J Food Sci* 76:C791-C794.
78. Chatterjee S, Sankaranarayanan R, Sonti RV. 2003. PhyA, a secreted protein of *Xanthomonas oryzae* pv. *oryzae*, is required for optimum virulence and growth on phytic acid as a sole phosphate source. *Mol Plant Microbe Interact* 16:973-982.
79. Yanke LJ, Bae HD, Selinger LB, Cheng KJ. 1998. Phytase activity of anaerobic ruminal bacteria. *Microbiology* 144:1565-1573.
80. Chu H-M, Guo R-T, Lin T-W, Chou C-C, Shr H-L, Lai H-L, Tang T-Y, Cheng K-J, Selinger BL, Wang AHJ. 2004. Structures of *Selenomonas ruminantium* phytase in complex with persulfated phytate: DSP phytase fold and mechanism for sequential substrate hydrolysis. *Structure* 12:2015-2024.
81. Puhl AA, Gruninger RJ, Greiner R, Janzen TW, Mosimann SC, Selinger LB. 2007. Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like *myo*-inositol polyphosphatase. *Protein Sci* 16:1368-1378.
82. Puhl AA, Greiner R, Selinger LB. 2008. A protein tyrosine phosphatase-like inositol polyphosphatase from *Selenomonas ruminantium* subsp. *lactilytica* has specificity for the 5-phosphate of *myo*-inositol hexakisphosphate. *Int J Biochem Cell Biol* 40:2053-2064.
83. Puhl AA, Greiner R, Selinger LB. 2008. Kinetics, substrate specificity, and stereospecificity of two new protein tyrosine phosphatase-like inositol polyphosphatases from *Selenomonas lactificex*. *Biochem Cell Biol* 86:322-330.

84. Zhou G, Denu JM, Wu L, Dixon JE. 1994. The catalytic role of Cys124 in the dual specificity phosphatase VHR. *J Biol Chem* 269:28084-28090.
85. Cirri P, Chiarugi P, Camici G, Manao G, Raugei G, Cappugi G, Ramponi G. 1993. The role of Cys12, Cys17 and Arg18 in the catalytic mechanism of low-Mr cytosolic phosphotyrosine protein phosphatase. *Eur J Biochem* 214:647-657.
86. Denu JM, Dixon JE. 1998. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol* 2:633-641.
87. Chen C-C, Cheng K-J, Ko T-P, Guo R-T. 2015. Current progresses in phytase research: three-dimensional structure and protein engineering. *ChemBioEng Reviews* 2:76-86.
88. Zhang ZY, Clemens JC, Schubert HL, Stuckey JA, Fischer MW, Hume DM, Saper MA, Dixon JE. 1992. Expression, purification, and physicochemical characterization of a recombinant *Yersinia* protein tyrosine phosphatase. *J Biol Chem* 267:23759-23766.
89. Beresford N, Patel S, Armstrong J, Szöör B, Fordham-Skelton Anthony P, Tabernero L. 2007. MptpB, a virulence factor from *Mycobacterium tuberculosis*, exhibits triple-specificity phosphatase activity. *Biochem J* 406:13-18.
90. Tiganis T, Bennett Anton M. 2007. Protein tyrosine phosphatase function: the substrate perspective. *Biochem J* 402:1-15.
91. Beresford NJ, Saville C, Bennett HJ, Roberts IS, Tabernero L. 2010. A new family of phosphoinositide phosphatases in microorganisms: identification and biochemical analysis. *BMC Genomics* 11:457.
92. Blüher D, Laha D, Thieme S, Hofer A, Eschen-Lippold L, Masch A, Balcke G, Pavlovic I, Nagel O, Schonsky A, Hinkelmann R, Worner J, Parvin N, Greiner R, Weber S, Tissier A, Schutkowski M, Lee J, Jessen H, Schaaf G, Bonas U. 2017. A 1-phytase type III effector interferes with plant hormone signaling. *Nat Commun* 8:2159.
93. Schenk G, Mitić N, Hanson GR, Comba P. 2013. Purple acid phosphatase: A journey into the function and mechanism of a colorful enzyme. *Coordin Chem Rev* 257:473-482.
94. Oddie GW, Schenk G, Angel NZ, Walsh N, Guddat LW, de Jersey J, Cassady AI, Hamilton SE, Hume DA. 2000. Structure, function, and regulation of tartrate-resistant acid phosphatase. *Bone* 27:575-584.
95. Kaija H, Alatalo SL, Halleen JM, Lindqvist Y, Schneider G, Kalervo Väänänen H, Vihko P. 2002. Phosphatase and oxygen radical-generating activities of mammalian purple acid phosphatase are functionally independent. *Biochem Biophys Res Commun* 292:128-132.
96. Bozzo GG, Raghothama KG, Plaxton WC. 2002. Purification and characterization of two secreted purple acid phosphatase isozymes from phosphate-starved tomato (*Lycopersicon esculentum*) cell cultures. *Eur J Biochem* 269:6278-6286.
97. Del Pozo JC, Allona I, Rubio V, Leyva A, De La Peña A, Aragoncillo C, Paz-Ares J. 1999. A type 5 acid phosphatase gene from *Arabidopsis thaliana* is induced by phosphate starvation and by some other types of phosphate mobilising/oxidative stress conditions. *Plant J* 19:579-589.
98. Veljanovski V, Vanderbeld B, Knowles VL, Snedden WA, Plaxton WC. 2006. Biochemical and molecular characterization of AtPAP26, a vacuolar purple acid phosphatase up-regulated in phosphate-deprived *Arabidopsis* suspension cells and seedlings. *Plant Physiol* 142:1282-1293.
99. Hegeman CE, Grabau EA. 2001. A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *Plant Physiol* 126:1598-1608.
100. Castillo Villamizar GA, Nacke H, Boehning M, Herz K, Daniel R. 2019. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio* 10:e01966-01918.
101. Ghorbani Nasrabadi R, Greiner R, Yamchi A, Nourzadeh Roshan E. 2018. A novel purple acid phytase from an earthworm cast bacterium. *J Sci Food Agric* 98:3667-3674.
102. Cangussu ASR, Aires Almeida D, Aguiar RWdS, Bordignon-Junior SE, Viana KF, Barbosa LCB, Cangussu EWdS, Brandi IV, Portella ACF, Dos Santos GR, Sobrinho EM, Lima WJN. 2018. Characterization of the catalytic structure of plant phytase, protein tyrosine phosphatase-like

- phytase, and histidine acid phytases and their biotechnological applications. *Enzyme Res* 2018:8240698-8240698.
103. Holme IB, Dionisio G, Madsen CK, Brinch-Pedersen H. 2017. Barley HvPAPhy_a as transgene provides high and stable phytase activities in mature barley straw and in grains. *Plant Biotechnol J* 15:415-422.
 104. Kong Y, Li X, Wang B, Li W, Du H, Zhang C. 2018. The soybean purple acid phosphatase GmPAP14 predominantly enhances external phytate utilization in plants. *Front Plant Sci* 9:292-292.
 105. Suzuki U, Yoshimura K, Takaishi M. 1907. Über ein enzym 'Phytase' das anhydro-oxy-methylen diphosphorsäure' spalter. *Tokyo Imperial University College Agricultural Bulletin* 7:503-512.
 106. Greiner R. 2007. Phytate-degrading enzymes: Regulation of synthesis in microorganisms and plants, p 78-96. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.
 107. Rodehutsord M, Rosenfelder P. 2016. Update on phytate degradation pattern in the gastrointestinal tract of pigs and broiler chickens, p 15-32. *In* Walk CL, Kühn I, Stein HH, Kidd MT, Rodehutsord M (ed), *Phytate destruction - consequences for precision animal nutrition* doi:10.3920/978-90-8686-836-0_1. Wageningen Academic, Wageningen, The Netherlands.
 108. Madhavan A, Sindhu R, Parameswaran B, Sukumaran RK, Pandey A. 2017. Metagenome analysis: A powerful tool for enzyme bioprospecting. *Appl Biochem Biotechnol* 183:636-651.
 109. Castillo Villamizar GA, Nacke H, Griese L, Tabernero L, Funkner K, Daniel R. 2019. Characteristics of the first protein tyrosine phosphatase with phytase activity from a soil metagenome. *Genes* 10:101.
 110. Farias N, Almeida I, Meneses C. 2018. New bacterial phytase through metagenomic prospection. *Molecules* 23:448.
 111. Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R. 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnol Lett* 34:663-675.
 112. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol Ecol* 78:188-201.
 113. Dukunde A, Schneider D, Lu M, Brady S, Daniel R. 2017. A novel, versatile family IV carboxylesterase exhibits high stability and activity in a broad pH spectrum. *Biotechnol Lett* 39:577-587.
 114. Sharma S, Khan FG, Qazi GN. 2010. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. *Appl Microbiol Biotechnol* 86:1821-1828.
 115. Waschkowitz T, Rockstroh S, Daniel R. 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol* 75:2506-2516.
 116. Daniel R. 2004. The soil metagenome – a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15:199-204.
 117. Lee MH, Lee S-W. 2013. Bioprospecting potential of the soil metagenome: novel enzymes and bioactivities. *Genomics Inform* 11:114-120.
 118. Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson AD, O'Gara F. 2014. Identification of novel phytase genes from an agricultural soil-derived metagenome. *J Microbiol Biotechnol* 24:113-118.

2**Function-Based Metagenomic Library Screening and Heterologous Expression Strategy for Genes Encoding Phosphatase Activity**

Genis A. Castillo Villamizar¹, Heiko Nacke¹, and Rolf Daniel^{1,2}

Methods Mol Biol. (2017), 1539:249-260

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen

Author contributions to the work:

Performed the experiments: GAC.

Analyzed data: GAC, HN.

Wrote the paper: GAC, HN, RD.

Conceived and designed the experiments: HN RD.

Function-Based Metagenomic Library Screening and Heterologous Expression Strategy for Genes Encoding Phosphatase Activity

Genis A. Castillo Villamizar, Heiko Nacke, and Rolf Daniel

Abstract

The release of phosphate from inorganic and organic phosphorus compounds can be mediated enzymatically. Phosphate-releasing enzymes, comprising acid and alkaline phosphatases, are recognized as useful biocatalysts in applications such as plant and animal nutrition, bioremediation and diagnostic analysis. Metagenomic approaches provide access to novel phosphatase-encoding genes. Here, we describe a function-based screening approach for rapid identification of genes conferring phosphatase activity from small-insert and large-insert metagenomic libraries derived from various environments. This approach bears the potential for discovery of entirely novel phosphatase families or subfamilies and members of known enzyme classes hydrolyzing phosphomonoester bonds such as phytases. In addition, we provide a strategy for efficient heterologous phosphatase gene expression.

Key words Phosphatases, Phytases, Metagenomic libraries, Phosphorus, Function-based screening

1 Introduction

Phosphorus is essential for growth, metabolism, and reproduction [1]. Due to enhanced demand and fertilization of agricultural land for food and biofuel production the consumption of phosphorus increased significantly during the last century. However, phosphate rock reservoirs renew in time scales of thousands to millions of years. Thus, mineral phosphorus resources are limited or will be even exhausted within the next 50–100 years [2]. Phosphorus is abundant in soil but present in its insoluble form or bound to organic compounds [3]. Consequently, this has led to the exploration of alternatives for obtaining phosphorus. The release of phosphorus in the form of phosphate can be mediated by a diverse group of enzymes. These enzymes, designated phosphatases, are considered as important biocatalysts for efficient phosphorus solubilization and release [4, 5]. Phosphatases showing phytase activity (phytases) are used to release phosphate from

phytate, the most abundant organic phosphorus compound in soil [6]. The released phosphate can then be utilized by for example agricultural crops as natural phosphorus fertilizer. Phosphatases catalyzing the hydrolysis of phytic acid also play an important role as supplement in animal nutrition, as they release phosphate from phytate present in cereal grains and oilseeds. Furthermore, phosphatases have broad applications in pharmaceutical industry and clinical diagnostics [7].

The almost exclusive use of cultivable microorganisms was a limiting factor with respect to the discovery of new enzymes exhibiting phosphatase activity. Within many recent culture-based approaches, degenerated primers were used to identify phosphatase genes carried by single microorganisms [8–10]. Taking into account that currently less than 1% of microbial taxa can be cultured under laboratory conditions, only a tiny fraction of the existing phosphatase gene pool has been mined by culture-based methods [8, 11–14]. In principle, culture-independent metagenomic approaches provide access to the entire phosphatase gene pool. In this way, novel phosphatases with valuable characteristics such as high stability and catalytic activity under harsh conditions can be identified. The different phosphatase types exhibit a high level of sequence divergence, and different substrate preferences and spectra [15]. These differences point to the employment of function-based screening strategies for the discovery of novel phosphatase-encoding genes from complex metagenomic libraries [11, 13, 16]. In contrast to sequence-based identification of target genes based on conserved DNA regions, the function-based strategy allows identification of enzymes that represent entirely novel phosphatase families.

Here, we describe a rapid function-based metagenomic library screening approach for phosphatase genes, which is based on chromogenic substrate-containing medium. Sarikhani and colleagues [11] used a similar medium for the discovery of phosphatase-encoding genes derived from *Pseudomonas putida*. We successfully tested phytic acid as well as other phosphorus sources such as β -glycerol phosphate in function-based screens of small- and large-insert metagenomic libraries. The number of retrieved phosphatase genes can vary depending on the source of environmental DNA used for the construction of metagenomic libraries. In addition, we present a strategy for efficient heterologous expression of genes conferring phosphatase activity. This strategy allows a moderate instead of a high level heterologous gene expression and the periplasmic localization of the heterologous phosphatase gene products. In this way the risk of detrimental interactions of the produced proteins with host proteins or cell metabolites in the cytoplasm is reduced.

2 Materials

2.1 Identification of Phosphatase Genes by Function-Based Screening of Metagenomic Libraries

2.1.1 Metagenomic Libraries

The function-based screening approach presented here has been tested using small-insert and large-insert metagenomic libraries derived from soil, compost, volcano sediments, glacial samples, and microbial mats. Metagenomic library construction was performed according to protocols described by Simon and Daniel [17]. Small-insert metagenomic libraries were constructed using the plasmid pCR-XL-TOPO (Thermo Fisher Scientific, Waltham, MA, USA) as vector. Large-insert metagenomic libraries were generated using the fosmid vector pCCIFOS™ (Epicentre Biotechnology, Madison, WI, USA).

2.1.2 Medium for Small-Insert Metagenomic Library Screening

1. Modified Sperber medium (SpM): 16 g/L agar, 10 g/L glucose or 2% glycerol, 500 mg/L yeast extract, 100 mg/L CaCl₂, and 250 mg/L MgSO₄, supplemented with a phosphorus source such as 2.5 g/L phytic acid, β-glycerol phosphate disodium salt pentahydrate, or D-fructose 6-phosphate disodium salt hydrate, 5-bromo-4-chloro-3-indolyl phosphate (BCIP) stock solution: 25 mg/mL in dimethylformamide.
2. Kanamycin stock solution: 50 mg/mL in H₂O.
3. NaOH solutions for pH adjustment.

2.1.3 Medium for Large-Insert Metagenomic Library Screening

The materials listed for small-insert metagenomic library screening medium can be used by considering the following modifications and extensions:

1. Chloramphenicol instead of kanamycin stock solution: 12.5 mg/mL in ethanol.
2. L-arabinose stock solution: 1% (w/v) in H₂O.

2.1.4 Function-Based Metagenomic Library Screening

1. *E. coli* DH5 alpha electrocompetent cells [18].
2. Super Optimal Broth with Catabolic repressor (SOC).
3. Bio-Rad GenePulser II (Bio-Rad, Munich, Germany) and 1 mm electroporation cuvettes.
4. Heat block with mixing function.

2.1.5 Analysis of Metagenomic DNA Fragments Carried by Positive Clones

1. Lysogeny broth (LB) (autoclaved).
2. Stock solution of antibiotic (kanamycin: 50 mg/mL, when pCR-XL-TOPO is used as vector or chloramphenicol: 12.5 mg/mL, when pCCIFOS is used as vector).
3. L-arabinose stock: 1% (w/v) or CopyControl™ Induction Solution (1000×, Epicentre Biotechnology) (required when pCCIFOS is used as vector).

4. **Plasmid** mini prep kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany).
5. *Hind*III restriction enzyme.
6. Sequencing primers: pCR-XL-TOPO vector: Forward 5'-GTAAAACGACGGCCAG-3', Reverse 5'-CAGGAAACAGCTATGAC-3', pCC1FOS: Forward 5'-GGATGTGCTGCAAGGCGATTAAGTTGG-3', Reverse 5'-CTCGTATGTGTGTGGAATTGTGAGC-3' (other appropriate primers can also be used).

2.2 Heterologous Expression of Phosphatase Genes

2.2.1 Cloning of Putative Phosphatase-Encoding Genes into Expression Vector

1. Phusion High-Fidelity DNA Polymerase PCR kit (Thermo Fisher Scientific GmbH, Schwerte, Germany) (other polymerases with proofreading activity can also be used).
2. pET-20b(+) Novagen vector (Merck KGaA, Darmstadt, Germany).
3. Gel extraction kit (Qiagen, Hilden, Germany).
4. Antarctic phosphatase (New England Biolabs GmbH, Frankfurt am Main, Germany).
5. DNA ligation kit (Thermo Fisher Scientific GmbH, Schwerte, Germany).
6. *E. coli* DH5 alpha electrocompetent cells.
7. Bio-Rad GenePulser II (Bio-Rad) and 1 mm electroporation cuvettes.
8. LB agar plates supplemented with 100 mg/L ampicillin.
9. LB broth (autoclaved).
10. Ampicillin stock solution: 100 mg/mL H₂O.

2.2.2 Heterologous Expression of Genes Encoding Phosphatase Activity

1. *E. coli* BL21 one shot cells (Thermo Fisher Scientific GmbH, Schwerte, Germany).
2. LB agar plates supplemented with 100 mg/L ampicillin.
3. Minimal medium is prepared from stock solution of 5× salts solution (50 g Na₂HPO₄·7H₂O, 30 g KH₂PO₄, 5 g NaCl, 5 g NH₄Cl) in 1 L water. For 1 L media, add 200 mL of the 10× salt solution to 500 mL of water supplemented with 0.20% glycerol, adjust the volume and sterilize by autoclaving. Supplement the media by adding 1 mL of MgSO₄ (1 M), 1 mL CaCl₂ (1 M), and 1 mL FeSO₄·7H₂O (0.01 mM). The solutions should be sterilized separately by filtration.
4. 50 mM Hepes buffer pH 8.
5. Shaker with temperature control.
6. French press or any other effective cell disruption device.

2.2.3 Verification of Heterologous Target Gene Expression Based on Phosphatase Activity Test

1. 50 mM sodium acetate buffer (pH 6.0).
2. Acetone.
3. 5 N H₂SO₄.
4. 10 mM ammonium molybdate.
5. 1 M citric acid.

3 Methods

3.1 Identification of Phosphatase Genes by Function-Based Screening of Metagenomic Libraries

The screening approach presented here is based on a chromogenic substrate (BCIP)-containing medium. This screening medium allows rapid identification of phosphatase activity encoded by small-insert and large-insert metagenomic libraries. We employed metagenomic libraries, which were constructed using plasmids or fosmids as vectors. In general, the described screening can also be performed using cosmid-based or bacterial artificial chromosome-based metagenomic libraries with modifications (e.g., use of appropriate antibiotics). Positive clones carrying potential phosphatase genes and exhibiting phosphatase activity show intense blue color after incubation on solidified screening medium. Sequencing and analysis of the metagenomic inserts derived from the isolated recombinant vectors of the positive clones allow prediction and identification of candidate genes responsible for the detected phosphatase activity.

3.1.1 Preparation of Screening Medium

1. To prepare screening medium, 500 mg/L yeast extract, 100 mg/L CaCl₂, 250 mg/L MgSO₄, and 2.5 g/L selected phosphorus source (*see Note 1*) are solubilized in water. Subsequently, pH is adjusted to 7.2 using NaOH solutions. To solidify the medium 16 g/L agar is added.
2. Autoclave the prepared mixture. After removing from the autoclave, allow cooling to approximately 55 °C and add filter-sterilized (0.22 µm filter) glucose (final concentration: 10 g/L) or autoclaved glycerol solution (final concentration: 2%). With respect to large-insert metagenomic library screening medium, filter-sterilized L-arabinose (0.22 µm filter) in a final concentration of 0.001 % or copy control induction solution is also added (*see Note 2*).
3. Add 1 mL/L of 25 mg/mL BCIP solution and appropriate antibiotic (kanamycin, final concentration: 50 mg/L or chloramphenicol, final concentration: 12.5 mg/L) to select for clones bearing small-insert or large-insert metagenomic libraries.
4. Pour the media into petri dishes. After solidification store at 4 °C and dark until use. Plates can be stored for up to 1 month under these conditions.

3.1.2 Function-Based Metagenomic Library Screening

1. Prechill the electroporation cuvette on ice.
2. Thaw *E. coli* DH5 alpha electrocompetent cells on ice and transfer 40 μL to the 1 mm electroporation cuvette.
3. Add 1 μL prepared metagenomic library DNA (DNA concentration approximately 350 ng/ μL) and mix gently. Do not mix by pipetting the cells up and down.
4. Wipe electrodes on the outside of cuvette with a paper towel to remove condensate and carefully eliminate air bubbles.
5. Electroporate the cells. We use a Bio-Rad Gene Pulser II with the following settings: 25 μF , 200 Ω and 1.25 kV.
6. Immediately add 500 μL of room temperature SOC medium.
7. Transfer the mix into a sterile 2 mL microcentrifuge tube and shake for 60 min at 37 °C and 150 rpm.
8. Spread 100 μL undiluted as well as diluted (tenfold, 100-fold, and 1000-fold) transformed cell suspension on separate screening medium plates containing the appropriate antibiotic. Store the remaining suspension of transformed cells at 4 °C.
9. Incubate the plates at 37 °C overnight. Analyze number of colonies formed on screening medium plates (*see Note 3*).
10. Spread the appropriate remaining undiluted or diluted suspension of transformed cells onto screening medium plates containing an appropriate antibiotic to obtain a sufficient number of colonies for detection of target clones.
11. Positive clones will appear after 24–72 h of incubation and show intense blue colony color resulting from reaction of phosphatase with the indicator BCIP (*see Note 4*).

3.1.3 Analysis of Metagenomic DNA Fragments Carried by Positive Clones

1. Pick single positive colonies and grow them individually in 5 mL LB broth (*see Note 2*) supplemented with the appropriate antibiotic (kanamycin, final concentration: 50 mg/L or chloramphenicol, final concentration: 12.5 mg/L).
2. Shake overnight at 37 °C and 150 rpm.
3. Extract, digest with restriction endonucleases, e.g., *Hind*III or any other enzyme present in the used vector and analyze insert DNA by using standard techniques.
4. Determine the insert sequences of vector DNA extracted from positive clones.
5. After the insert DNA sequences have been determined, identify open reading frames (ORFs). An initial prediction of ORFs can be performed using the ORF finder tool provided by the National Center for Biotechnology Information [19, 20].
6. To identify ORFs potentially conferring phosphatase activity, examine coding sequences for similarities to protein families

and domains, e.g., by performing searches against the CDD databases [21]. Consider that the described function-based screening approach allows identification of members of previously unknown phosphatase families. In some cases, the similarity of identified ORFs to known phosphatase sequences might be very low.

3.2 Efficient Heterologous Expression of Phosphatase Genes

Enzymes hydrolyzing phosphomonoester bonds play an important role in regulation of cell metabolism. The heterologous expression strategy described here has been developed to minimize interactions of recombinant phosphatases with host cell molecular activities and putative toxic effects. Metagenomic library-derived target genes are cloned into an expression vector encoding a signal sequence for periplasmic localization to reduce reactions of recombinant phosphatases with biomolecules in the cytoplasm of the host cell. To further reduce potential detrimental effects of phosphatase activity, we recommend conditions that lead to a moderate instead of a high level heterologous gene expression. These conditions include e.g., the use of an appropriate minimal medium rather than complex medium during heterologous phosphatase gene expression and protein production. To verify heterologous production of the enzyme, a phosphatase activity assay should be performed.

3.2.1 Cloning of Putative Phosphatase-Encoding Genes into Expression Vector

1. Design primers for amplification of a putative phosphatase gene. In order to clone the gene into expression vector pET-20b(+), add restriction sites occurring in the multiple cloning site (MCS) of this vector to the primers (one restriction site per primer). To allow directional cloning, each primer should contain a different restriction site. Within the MCS, the different selected restriction sites should be separated by at least 10 bp. Ensure that the selected restriction sites are not present in the gene region encoding the putative phosphatase.
2. Check if the designed primers allow cloning of the PCR product in frame with the His₆ tag and the signal sequence for periplasmic localization encoded by plasmid pET-20b(+).
3. Perform a PCR to amplify the putative phosphatase gene using the primers containing the added restriction sites. We use Phusion High Fidelity Hot Start DNA polymerase to obtain PCR products. The PCR reaction mixture (50 μ L) contains 10 μ L of fivefold Phusion GC buffer, 200 μ M of each dNTP, 1.5 mM MgCl₂, 2 μ M of each of the primers, 2.5% DMSO, 0.5 U Phusion High Fidelity Hot Start DNA polymerase (*see Note 5*), and approximately 25 ng recombinant plasmid or fosmid DNA extracted from a positive clone.
4. The thermal cycling scheme should be adjusted by considering the size of potential phosphatase genes and annealing temperature of selected primers. A gradient PCR is recommended to

quickly identify an appropriate annealing temperature. The following thermal cycling scheme can be used to test different annealing temperatures: initial denaturation at 98 °C for 2 min, 29 cycles of denaturation at 98 °C for 45 s, annealing gradient ranging from 58 to 68 °C for 45 s, and extension at 72 °C for 30 s per kb, followed by a final extension at 72 °C for 5 min. Check the PCR products by agarose gel electrophoresis. Further PCR reactions using the selected annealing temperature can be performed to obtain a higher amount of the PCR product.

5. Subject the PCR product to agarose gel electrophoresis (0.8%) and purify it using a gel extraction kit, e.g., QIAquick (Qiagen, Hilden, Germany).
6. **Digest** the PCR product and pET-20b(+) vector separately using the restriction enzymes selected for directional cloning. Due to the loss of DNA during the following gel purification step, it is important to digest at least 1 µg PCR product and 3 µg pET-20b(+) vector.
7. Dephosphorylation of pET-20b(+) vector. In order to prevent re-circularization of the vector pET-20b(+), the digested plasmid should be treated with a phosphatase prior to the following gel purification step. Add a maximum of 2 U Antarctic phosphatase (1 U/µL) to the pET-20b(+) vector restriction digest. The Antarctic phosphatase is stable and active in most restriction digestion buffers. Incubate for 15 min at 37 °C.
8. Load the digested PCR product and plasmid DNA separately on a 0.8% agarose gel and purify both using a gel extraction kit.
9. Ligation: the ligation mix (20 µL) contains 2 µL 100 mM DTT, 1 µL 10 mM ATP, 2 µL 10× T4 Ligase buffer, approximately 200 ng pET-20b(+) (digested, dephosphorylated and purified), PCR product (digested and purified), and 1 µL T4 ligase (1 U/µL). A molar ratio 1:3 of vector to PCR product is recommended. Incubate at 16 °C overnight. In order to improve the ligation efficiency the T4 ligase is subsequently inactivated by incubation at 65 °C for 10 min.
10. Transform *E. coli* DH5 alpha electrocompetent cells with 5 µL ligation reaction via electroporation. Spread diluted (tenfold and 100-fold) suspension of the transformed cells on LB plates containing ampicillin and incubate overnight at 37 °C. Pick six single colonies and grow them in 5 mL LB broth with ampicillin (final concentration: 100 mg/L) overnight at 37 °C and 150 rpm. Extract the plasmid DNA using standard techniques.
11. Digest extracted plasmids, using the restriction enzymes selected for directional cloning, and check for the presence of the insert by agarose gel electrophoresis.

12. Sequence plasmid DNA carrying the desired insert to verify that the putative phosphatase gene has been cloned in the correct orientation and the sequence is error-free.

3.2.2 Heterologous Expression of Genes Encoding Phosphatase Activity

1. Transform *E. coli* BL21 one shot cells. Thaw one tube of *E. coli* BL21 one shot cells on ice and subsequently add 1 μ L of the constructed expression plasmid harboring the target gene (maximum of 30 ng). Incubate on ice for 30 min. Perform transformation of the recombinant plasmids into the cells by heat shock treatment at 42 °C for 30 s in a temperature-controlled water bath. Immediately transfer the tube to ice and subsequently add 250 μ L SOC medium. Incubate the tube at 37 °C for 45 min. Spread 100 as well as 150 μ L suspension of transformed cells on LB plates containing ampicillin (*see Note 6*). Incubate plates overnight at 37 °C.
2. Pick 3–4 colonies and grow each in 30 mL minimal medium containing ampicillin (final concentration: 100 mg/L) overnight at 30 °C and 150 rpm.
3. Use the overnight culture to inoculate 250 mL of minimal medium (resulting OD₆₀₀ should be approximately 0.1). Incubate the culture with shaking (150 rpm) at 30 °C until it reaches log phase (OD₆₀₀ 0.4–0.8). Induce the production of the recombinant protein by adding IPTG to a final concentration of 0.25 mM and incubate with shaking (150 rpm) at 30 °C until OD₆₀₀ of approximately 3.2 (*see Note 7*).
4. Harvest the cells by centrifugation at 10,000 $\times g$ and 4 °C for 20 min. Suspend the resulting cell pellet in chilled lysis buffer 50 mM HEPES pH 7.5. Use a ratio of 1:2 w/v of pellet to buffer (*see Note 8*). Disrupt the cells using a prechilled French Press cell (1.38 $\times 10^8$ Pa).
5. Clarify the cell lysate by centrifugation for 20 min at 9000 $\times g$ and 4 °C. The supernatant (crude extract) should be cleared by filtration using a 0.2 μ m syringe filter. *Note:* Subsequent purification methods of the crude extract can be applied to purify the target protein but a check for phosphatase activity should be performed using the crude extract.

3.2.3 Verification of Heterologous Target Gene Expression Based on Phosphatase Activity Test

1. To identify activity of target proteins, released inorganic phosphate can be measured according to the ammonium molybdate method [22] with modifications. Add 10 μ L of diluted crude extract or purified enzyme to 350 μ L of 50 mM sodium acetate buffer (pH 5.0) and incubate for 3 min at 40 °C. Add 10 μ L of 100 mM phosphorus source used for function-based screening (some commercial substrates contain traces of free phosphorus causing background coloring). For blanks, use 10 μ L lysis buffer (for purified potential phosphatase) or crude extract derived from *E. coli* BL21 carrying empty pET-20b(+) vector (for non-purified samples).

2. After 30 min of incubation at 40 °C, add 1.5 mL of a freshly prepared solution of acetone/5 N H₂SO₄/10 mM ammonium molybdate (2:1:1 v/v) and 100 µL of 1 M citric acid. All assays should be performed in triplicate. When phosphorus has been released in presence of molybdate, a bright yellow phosphomolybdate complex is formed and extracted by acetone. It is possible that the yellow color is directly visible. However, a spectrophotometer measurement at 355 nm is recommended [22].

4 Notes

1. We successfully tested all three phosphorus sources phytic acid, β-glycerol phosphate disodium salt pentahydrate, and D-fructose 6-phosphate disodium salt hydrate with respect to function-based identification of phosphatase genes from small-insert and large-insert metagenomic libraries. However, phosphorus source selection will depend on the research approach and target group of phosphatases. For example, to increase probability of identifying phytases during function-driven screens the phosphorus source phytic acid should be selected as screening substrate. In the case of metagenomic library screens this substrate might also act as inducer for expression of genes encoding phytases via endogenous promoters [23].
2. The number of fosmid carried by *E. coli* clones can be increased by adding L-arabinose (final concentration: 0.001%) or CopyControl™ Induction Solution (1000×). This might be advantageous during screening of large-insert metagenomic libraries as increased copy numbers of fosmid containing target genes might result in an increase of total phosphatase activity. Furthermore, higher DNA amounts can be extracted from clones carrying multiple copies of fosmid compared to those harboring a single fosmid.
3. To facilitate identification and selection of individual positive clones, we recommend growing a maximum number of approximately 10,000 colonies on screening medium plates (petri dishes 150 × 20 mm).
4. Due to endogenous phosphatase activity of host cells, all colonies grown on screening medium plates will change color after prolonged incubation times (more than 48 h). Positive clones show intense blue colony color, whereas false-positive colonies exhibit light blue or greenish color. Thus, it is possible that in some cases the presence of weakly expressed metagenome-derived phosphatases showing low catalytic activity is masked by the background reaction of the screening host. Nevertheless, we were able to identify a high number of clones carrying

recombinant phosphatase genes based on intense blue color developed by individual *E. coli* colonies.

5. It is highly recommended to use a proofreading polymerase to minimize mutations during amplification of putative phosphatase genes.
6. It is possible to spread some of the transformed cells directly on Sperber medium plates containing BCIP and ampicillin. Intense blue colonies growing on these plates indicate the presence of the targeted phosphatase gene.
7. The described conditions for heterologous expression were successfully tested for a number of phosphatase genes derived from function-based metagenomic library screening. Nevertheless, it might be necessary to vary different parameters such as temperature, IPTG concentration, and incubation time for improving heterologous expression of individual phosphatase genes.
8. The lysis buffer should be modified depending on intended further analysis. For instance, for affinity chromatography using a Ni-column 300 mM NaCl can be added.

References

1. McDowell LR (2003) Chapter 2 - Calcium and phosphorus. In: McDowell LR (ed) Minerals in animal and human nutrition, 2nd edn. Elsevier, Amsterdam, pp 33–100
2. Smil V (2000) Phosphorus in the environment: natural flows and human interferences. *Annu Rev Energy Environ* 25:53–88
3. Tarafdar JC, Marschner H (1994) Phosphatase activity in the rhizosphere and hyphosphere of VA mycorrhizal wheat supplied with inorganic and organic phosphorus. *Soil Biol Biochem* 26:387–395
4. Bagyaraj DJ, Krishnaraj PU, Khanuja SPS (2000) Mineral phosphate solubilization: agronomic implications, mechanism and molecular genetics. *Proc Indian Nat Sci Acad B Rev Tracts Biol Sci* 66:69–82
5. Cromwell GL (2009) ASAS Centennial Paper: landmark discoveries in swine nutrition in the past century. *J Anim Sci* 87:778–792
6. Lim BL, Yeung P, Cheng C, Hill JE (2007) Distribution and diversity of phytate-mineralizing bacteria. *ISME J* 1:321–330
7. Muginova SV, Zhavoronkova AM, Polyakov AE, Shekhovtsova TN (2007) Application of alkaline phosphatases from different sources in pharmaceutical and clinical analysis for the determination of their cofactors; Zinc and Magnesium ions. *Anal Sci* 23:357–363
8. Greiner R (2004) Purification and Properties of a Phytate-degrading Enzyme from *Pantoea agglomerans*. *Protein J* 23:567–576
9. Cho J, Lee C, Kang S, Lee J, Lee H, Bok J et al (2005) Molecular cloning of a phytase gene (*phy M*) from *Pseudomonas syringae* MOK1. *Curr Microbiol* 51:11–15
10. Cheng W, Chiu CS, Guu YK, Tsai ST, Liu CH (2013) Expression of recombinant phytase of *Bacillus subtilis* E20 in *Escherichia coli* HMS 174 and improving the growth performance of white shrimp, *Litopenaeus vannamei*, juveniles by using phytase-pretreated soybean meal-containing diet. *Aquacult Nutr* 19:117–127
11. Sarikhani M, Malboobi M, Aliasgharadz N, Greiner R, Yakhchali B (2010) Functional screening of phosphatase-encoding genes from bacterial sources. *Iran J Biotech* 8:275–279
12. Riccio ML, Rossolini GM, Lombardi G, Chiesurin A, Satta G (1997) Expression cloning of different bacterial phosphatase-encoding genes by histochemical screening of genomic libraries onto an indicator medium containing phenolphthalein diphosphate and methyl green. *J Appl Microbiol* 82:177–185
13. Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson ADW, O’Gara F (2014) Identification of novel phytase genes from an

- agricultural soil-derived metagenome. *J Microbiol Biotechnol* 24:113–118
14. Yao MZ, Zhang YH, Lu WL, Hu MQ, Wang W, Liang AH (2012) Phytases: crystal structures, protein engineering and potential biotechnological applications. *J Appl Microbiol* 112:1–14
 15. Kennelly PJ (2001) Protein phosphatases – a phylogenetic perspective. *Chem Rev* 101:2291–2312
 16. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L et al (2015) Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc Natl Acad Sci U S A* 112:E1974–E1983
 17. Simon C, Daniel R (2010) Construction of small-insert and large-insert metagenomic libraries. *Methods Mol Biol* 668:39–50
 18. Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* 16:6127–6145
 19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
 20. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K et al (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 40:D13–D25
 21. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC et al (2013) CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 41:D348–D352
 22. Heinonen JK, Lahti RJ (1981) A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase. *Anal Biochem* 113:313–317
 23. Vijayaraghavan P, Primiya RR, Prakash Vincent SG (2013) Thermostable alkaline phytase from *Alcaligenes* sp. in improving bioavailability of phosphorus in animal feed: *in vitro* analysis. *ISRN Biotechnol* 2013:6

3**Functional Metagenomics Reveals an Overlooked Diversity and Novel Features of Soil-Derived Bacterial Phosphatases and Phytases**

Genis Andrés Castillo Villamizar^{1,2}, Heiko Nacke¹, Marc Boehning¹, Kristin Herz¹, Rolf Daniel¹

mBio (2019), 10:e01966-18

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany,

²Línea tecnológica biocorrosión, Corporación para la investigación de la corrosión C.I.C. Piedecuesta, Santander, Colombia

Author contributions to the work:

Performed the experiments: GAC, MB, KH.


Analyzed data: GAC, HN, RD.

Wrote the paper: GAC, HN, RD.

Developed the concept and designed the experiments: HN, RD.



Functional Metagenomics Reveals an Overlooked Diversity and Novel Features of Soil-Derived Bacterial Phosphatases and Phytases

Genis Andrés Castillo Villamizar,^{a,b} Heiko Nacke,^a Marc Boehning,^{a*} Kristin Herz,^a  Rolf Daniel^a

^aDepartment of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany

^bLínea tecnológica biocorrosión, Corporación para la investigación de la corrosión C.I.C. Piedecuesta, Santander, Colombia

ABSTRACT Phosphatases, including phytases, play a major role in cell metabolism, phosphorus cycle, biotechnology, and pathogenic processes. Nevertheless, their discovery by functional metagenomics is challenging. Here, soil metagenomic libraries were successfully screened for genes encoding phosphatase activity. In this context, we report the largest number and diversity of phosphatase genes derived from functional metagenome analysis. Two of the detected gene products carry domains which have never been associated with phosphatase activity before. One of these domains, the SNARE-associated domain DedA, harbors a so-far-overlooked motif present in numerous bacterial SNARE-associated proteins. Our analysis revealed a previously unreported phytase activity of the alkaline phosphatase and sulfatase superfamily (cl23718) and of purple acid phosphatases from nonvegetal origin. This suggests that the classical concept comprising four classes of phytases should be modified and indicates high performance of our screening method for retrieving novel types of phosphatases/phytases hidden in metagenomes of complex environments.

IMPORTANCE Phosphorus (P) is a key element involved in numerous cellular processes and essential to meet global food demand. Phosphatases play a major role in cell metabolism and contribute to control the release of P from phosphorylated organic compounds, including phytate. Apart from the relationship with pathogenesis and the enormous economic relevance, phosphatases/phytases are also important for reduction of phosphorus pollution. Almost all known functional phosphatases/phytases are derived from cultured individual microorganisms. We demonstrate here for the first time the potential of functional metagenomics to exploit the phosphatase/phytase pools hidden in environmental soil samples. The recovered diversity of phosphatases/phytases comprises new types and proteins exhibiting largely unknown characteristics, demonstrating the potential of the screening method for retrieving novel target enzymes. The insights gained into the unknown diversity of genes involved in the P cycle highlight the power of function-based metagenomic screening strategies to study Earth's phosphatase pools.

KEYWORDS SNARE-associated domain, functional metagenomics, phosphatases, phytases, soil metagenome

Within the last decades, advances in next-generation sequencing and metagenomic techniques have led to the discovery of new enzymes from metagenomes (1, 2). Novel lipases, esterases, proteases, and hydrogenases, among many others, have been identified (3, 4). Nevertheless, the majority of enzymes with high biological relevance are still almost exclusively recovered from cultured organisms (2).

Citation Castillo Villamizar GA, Nacke H, Boehning M, Herz K, Daniel R. 2019. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio* 10:e01966-18. <https://doi.org/10.1128/mBio.01966-18>.

Editor Daniel Barkan, Hebrew University of Jerusalem

Copyright © 2019 Castillo Villamizar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Rolf Daniel, rdaniel@gwdg.de.

* Present address: Marc Boehning, Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.

Received 6 September 2018

Accepted 11 December 2018

Published 29 January 2019

This is especially the case for phosphatases. Phosphatases have evolved across all living organisms and contribute to the regulation of diverse cellular functions (5, 6). A specific group of phosphatases named phytases can release phosphorus from phytic acid, which is one of the most important phosphorus reserves in plants and soils (7, 8).

Phosphorus (P) reserves are globally important, due to the enormous growth of the world population, and the ensuing demand for this macroelement. Large amounts of P are and will be required in order to fulfill the increasing world agroalimentary needs (9). However, global rock phosphorus reservoirs are currently being rapidly depleted, and the supplementation of P to animal feed and plant fertilizers has become more expensive during the last decades (10). Plant-based animal feeds often contain large amounts of phytate, which cannot be utilized by monogastric animals due to the lack of phytases (7, 11). As a consequence, P levels in soils and water bodies increase. This eutrophication causes for instance algal blooms in aquatic ecosystems, leading to deoxygenated areas disturbing the life of many species (12). To meet future requirements, minimize losses of P, and reduce the environmental impact, it is necessary to use P compounds more efficiently and develop economical recycling technologies. In this context, phosphatases/phytases have proved to be remarkably useful (13). These enzymes are currently used in agroindustry to minimize P losses and to improve the levels of bioavailable P (14). A more recently described role of the phytases is their involvement in pathogenicity causing tissue damage in humans, coordination of the virulence program in *Dickeya dadantii*, and mediation of plant infection by *Candida albicans* and *Xanthomonas*, respectively (5, 15, 16).

The diversity and potential of environmental phytases remain largely unexplored as so far almost all reported functionally characterized phytases were derived from cultured organisms, including plants, fungi, and bacteria. Based on their catalytic characteristics, four classes of phytases have been described: histidine acid phytase (HAPhy), β -propeller phytase (BPPhy), purple acid phytase (PAPhy), and protein tyrosine phytase (PTPhy). These enzymes are structurally and catalytically dissimilar (14, 17).

In this study, we use a function-based screening approach (18) to identify environmental phosphatases/phytases. By using soil metagenomes as a source, we were able to recover novel genes encoding phosphatases with phytase activity. Some of the recovered genes encode protein domains that were not associated with phosphatase activity before, and others represent new types or subtypes of phytases.

RESULTS

Phosphatase detection strategy. The metagenomic libraries contained approximately 38,122 to 166,040 clones and were screened for candidates exhibiting phosphatase activity using plates with phytate as phosphorus source and BCIP as indicator (see Fig. S1 in the supplemental material). The quality of the libraries was controlled by determining the average insert sizes and the percentage of insert-bearing *Escherichia coli* clones. The average insert sizes of metagenomic DNA-containing plasmids ranged from 2.8 to 6.7 kb, and the frequency of clones carrying plasmid inserts was at least 89% (Table 1).

We recovered 21 positive *E. coli* clones from functional screens carrying plasmids harboring one or more ORFs associated with known phosphatase genes and domains (designation of plasmids is given in Table 1). The entire inserts of the positive clones were sequenced and taxonomically classified, showing that in all cases the cloned environmental DNA is of bacterial origin. Most inserts of the positive clones were affiliated with *Terrabacteria*, *Proteobacteria*, and the PVC superphylum with seven, six, and four representatives, respectively. Within the *Terrabacteria* group, most of the inserts (4) were affiliated with *Actinobacteria* (Table S1).

Thirty-one ORFs encoding putative gene products with similarity to known phosphatase enzymes were identified. Signal peptides were detected for 12 of them. The deduced gene products comprised 214 to 819 amino acids with calculated molecular masses ranging from 12 to 65.5 kDa and amino acid sequence identities to the closest

TABLE 1 Characteristics of the soil metagenomic libraries and designation of plasmids harbored by positive clones

Library ^a	No. of clones	Avg insert size (kb)	Insert frequency (%)	Estimated library size (Gb)	No. of positive clones/Gb	Plasmid(s) of positive clones (accession no.)
AEW1*	129,748	6.7	91	0.79	1.2	pLP01 (KY931670)
AEW5*	90,300	5.2	89	0.42	2.3	pLP02 (KY931671)
SEW2*	135,240	5.7	95	0.73	9.6	pLP10 (KY931677), pLP14 to pLP19 (KY931679 to KY931684)
SEW5*	166,040	4.0	95	0.63	1.6	pLP07 (KY931674)
SEW46	38,122	2.8	93	0.17	23.5	pLP03 (KY931672), pLP04 (KY931673), pLP08 (KY931675), pLP09 (KY931676)
HEW30	53,460	6.1	96	0.31	22.6	pLP13 (KY931678), pLP20 (KY931685), pLP24 to pLP28 (KY931686 to KY931690)

^aAEW, metagenomic libraries derived from the Biodiversity Exploratory Schwäbische Alb; SEW, metagenomic libraries derived from the Biodiversity Exploratory Schorfheide-Chorin; HEW, metagenomic libraries derived from the Biodiversity Exploratory Hainich-Dün. *, previously generated libraries (39).

known phosphatases ranging from 25% (Pho14B) to 83% (Pho13) over the full-length protein (Table 2).

From the 21 positive clones, seven harbored more than one putative phosphatase-related gene (Table 2). Thus, if two or more potential phosphatase activity-related genes were present in a positive clone, individual heterologous expression and subsequent phosphatase activity verification were performed. The analysis of colonies showed that the individual heterologous expression of 24 out of 31 genes led to phosphatase activity and the corresponding positive phenotype of the respective recombinant *E. coli* strains (Table 2).

High phosphatase diversity recovered from soil metagenomes. Phosphatases can be classified according to the structural fold of the catalytic domains and subclassified into families and subfamilies based on sequence similarities of the phosphatase domains, as well as by conserved amino acid motifs not belonging to the catalytic domain (6, 19). However, some are still classified based on their biochemical properties and biological functions (20).

Among the putative gene products encoded by the 31 candidate genes, alkaline phosphatases were identified as the most abundant group (five representatives), followed by histidine phosphatases and phospholipases with four representatives each. Phosphoserine-phosphatases and protein-tyrosine phosphatases were represented by three putative genes each. Acid phosphatases were encoded by two genes, while the plasmid pLP10 harbored an ORF with a deduced gene product showing similarity to a mismatch repair ATPase (Table 2).

The amino acid sequence analysis revealed the presence of 10 different domains in the 31 deduced proteins. We detected the alkaline phosphatase and sulfatase superfamily domain (ALP-like cl23718) as the most frequent domain, represented in eight sequences. The second highest abundance showed the haloacid dehydrogenase domain (HAD cl21460), which was identified in six protein sequences. Three out of four classical phosphatase/phytase domains were detected in this study: the histidine phosphatase domain (HP with five protein sequences), the tyrosine phosphatase domain (PTPc with two protein sequences), and the acid phosphatase domain (PAP with two protein sequences) (Fig. 1). The phylogenetic analyses of the enzyme sequences and those harboring the above-mentioned domains revealed different clustering patterns in relation to reference phosphatase sequences for the different groups. Within the analyzed groups, the clustering of the metagenome-derived enzymes ranged from clear separation to integrated clustering (Fig. S2).

The HP superfamily (cl11399) is represented by a diverse group of proteins divided into two branches exhibiting numerous functions (21). Classical members of the HAPhy share a conserved motif, RHGXRX, characteristic for this enzyme class. The HAPhy catalytic reactions are based on the conserved histidine residue in the RHGXRX motif (21, 22). In this study, all five phosphatases belonging to the HP superfamily harbored this histidine residue (Fig. 2A). Three out of five HPs in this survey were encoded by

TABLE 2 Gene products encoded by genes associated with phosphatase activity and their observed sequence identities

Gene (accession no. of protein)	No. of amino acids encoded	Closest similar phosphatase protein, accession no. (no. of encoded amino acids), organism, E value	Identity to closest similar phosphatase protein (Blast), no. of amino acids similar/total no. (%)	% identity to closest similar phosphatase protein (Clustal alignment)
<i>pho01</i> (AWN00218)	229	Phosphatidylglycerophosphatase, PIF15492.1 (224), <i>Rhodanobacter</i> sp. strain TND4EH1, 3E−99	161/213 (76)	72
<i>pho02</i> (AWN00219) ^a	339	Phosphoserine phosphatase, AFM25187 (342), <i>Desulfomonile tiedjei</i> DSM 6799, 0.0	251/337 (74)	74
<i>pho03A</i> (AWN00220) ^a	493	Phosphoesterase, WP_009239878.1 (404), <i>Ralstonia</i> , 2E−9	183/425 (49)	47
<i>pho03B</i> (AWN00221) ^b	222	Phospholipase/carboxylesterase, ADV48687.1 (334), <i>Cellulophaga algicola</i> DSM 14237, 2E−14	84/181 (46)	27
<i>pho04</i> (AWN00222)	214	Putative membrane-associated alkaline phosphatase, KGB26473 (203), <i>Acetobacter tropicalis</i> , 9E−50	92/193 (48)	46
<i>pho07</i> (AWN00223) ^a	392	Phosphoesterase family protein, PZS03611.1 (379), <i>Pseudonocardiales</i> bacterium, 1E−111	184/349 (53)	51
<i>pho08A</i> (AWN00224) ^a	235	Histidine phosphatase family protein, WP_074262886.1 (229), <i>Paraburkholderia phenazinium</i> , 2E−56	97/191 (51)	49
<i>pho08B</i> (AWN00225) ^a	236	Histidine phosphatase family protein, WP_090546752.1 (196), <i>Paraburkholderia caballeronis</i> , 1E−59	97/182 (53)	50
<i>pho08C</i> (AWN00226) ^a	238	Histidine phosphatase family protein, WP_090546752.1 (196), <i>Paraburkholderia caballeronis</i> , 2E−57	98/182 (54)	51
<i>pho09C</i> (AWN00227) ^a	455	Alkaline phosphatase family protein, WP_007415052.1 (407), <i>Pedospaera parvula</i> , 0.0	330/413 (66)	63
<i>pho10</i> (AWN00228)	554	Mismatch repair ATPase, WP_014786775 (599), <i>Terriglobus roseus</i> , 6E−142	246/558 (44)	44
<i>pho13</i> (AWN00229)	411	Broad-specificity phosphatase PhoEn, WP_071949433.1 (401), <i>Mycobacterium</i> sp. strain PYR15, 0.0	349/400 (87)	83
<i>pho14A</i> (AWN00230) ^{a,b}	229	Protein tyrosine phosphatase (partial), CCZ50566.1 (64), <i>Acidobacteria</i> bacterium, 9E−13	43/111 (50)	48
<i>pho14B</i> (AWN00231) ^b	305	Phosphoserine phosphatase, PKM89459.1 (276), <i>Firmicutes</i> bacterium, 2E−4	58/215 (27)	25
<i>pho14C</i> (AWN00232) ^b	356	Phosphatidylserine/phosphatidyl glycerophosphate, AEQ20292 (371), uncultured bacterium CSLG7, 2E−109	175/357 (49)	48
<i>pho14D</i> (AWN00233) ^b	602	Protein tyrosine phosphatase, PYO70860.1 (581), <i>Gemmatimonadetes</i> bacterium, 1E−137	244/579 (42)	41
<i>pho15</i> (AWN00234)	223	Alkaline phosphatase, OFV86354.1 (209), <i>Acidobacteria</i> bacterium, 8E−34	71/167 (43)	41
<i>pho16A</i> (AWN00235)	819	Diguanylate cyclase/phosphodiesterase, WP_067501625.1 (816), <i>Actinoplanes</i> sp. strain TFC3, 1E−46	105/247 (43)	39
<i>pho16B</i> (AWN00236) ^a	376	Protein tyrosine phosphatase, WP_042381880.1 (372), <i>Streptacidiphilus melanogenes</i> , 0.0	257/324 (89)	76
<i>pho17A</i> (AWN00237) ^a	353	Phosphoserine phosphatase, AFM25187 (342), <i>Desulfomonile tiedjei</i> DSM 6799, 0.0	252/329 (77)	74
<i>pho18</i> (AWN00238) ^a	248	Phosphatase PAP2 family protein, WP_093286091.1 (257), <i>Verrucomicrobiaceae</i> bacterium GAS474, 4E−55	99/200 (50)	46
<i>pho19A</i> (AWN00239)	612	Alkaline phosphatase precursor, AMY11511 (577), <i>Acidobacteria</i> bacterium DSM 100886, 8E−126	230/529 (43)	42
<i>pho20B</i> (AWN00240)	392	Phosphoglycolate phosphatase, RDI59778.1 (337), <i>Microvirga subterranea</i> , 3E−152	248/339 (76)	73
<i>pho24</i> (AWN00241)	428	PAP2 superfamily protein, SHK15444 (414), <i>Bradyrhizobium lablabi</i> , 3E−141	215/405 (53)	54
<i>pho25B</i> (AWN00242) ^{a,b}	526	Phospholipase, WP_052891151 (505), <i>Thermogemmatispora carboxidivorans</i> , 0.0	303/527 (57)	60
<i>pho25C</i> (AWN00243)	252	Phospholipase, WP_006679394.1 (222), <i>Paenibacillus dendritiformis</i> , 0.0	41/101 (41)	28
<i>pho26</i> (AWN00244) ^a	559	Alkaline phosphatase family protein, WP_020714678.1 (564), <i>Acidobacteriaceae</i> bacterium KBS 89, 0.0	434/551 (79)	78
<i>pho27A</i> (AWN00245) ^a	347	Multispecies: phosphatase, WP_PYV87257.1 (338), <i>Acidobacteria</i> bacterium, 9E−64	249/323 (77)	74
<i>pho27B</i> (AWN00246)	263	Acid sugar phosphatase, GBD30013.1 (265), bacterium HR32, 2E−57	106/254 (42)	39
<i>pho28A</i> (AWN00247) ^b	490	Nonhemolytic phospholipase C, APW61637.1 (486), <i>Paludisphaera borealis</i> , 0.0	328/454 (72)	69
<i>pho28C</i> (AWN00248) ^a	232	Histidine phosphatase family protein, WP_106819986.1 (214), <i>Syntrophobacter</i> sp. strain SbD1, 2E−61	93/170 (53)	46

^aSignal peptide detected.^bNo phosphatase activity was detected on indicator plates after cloning ORF into expression vector.

plasmid pLP08. The analysis of the plasmid sequence revealed a tandem organization of these genes with slight individual sequence variations (Fig. 2A; Fig. S3).

PTPs are well-studied proteins with a characteristic motif (HCX5R) (23, 24). In this study, two new PTPs (Pho14A and Pho16B) harboring the typical catalytic signature of the group (Fig. 2B) were detected. Interestingly, Pho16B showed the specific signature

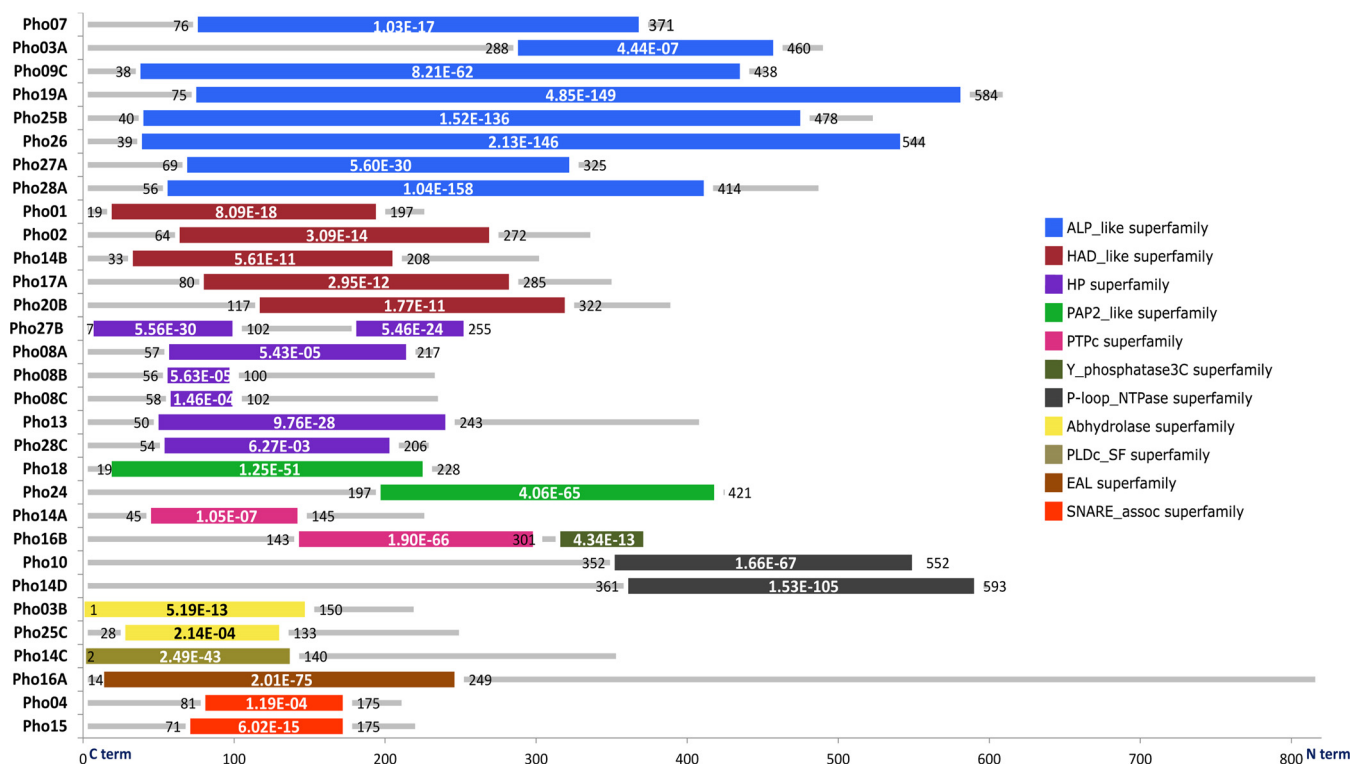


FIG 1 General architecture and domains of the retrieved phosphatases: ALP, alkaline phosphatases and sulfatases (cl23718); HAD, haloacid dehalogenase (cl21460); HP, histidine phosphatase (cl11399); PAP2, phosphatidic acid phosphatase (cl00474); PTPs, protein tyrosine phosphatases (cl21483); Y phosphatase 3C superfamily (cl6249); P-loop NTPase superfamily (cl21455); abhydrolase superfamily (cl21494); PLDc, phospholipase D (cl15239); EAL superfamily (cl00290); SNARE-associated superfamily (cl00429).

of the MtpB-like phosphatases characterized by the presence of the unique active site P-loop submotif HCXXKDRDRT. This type of protein has been predicted in several microorganisms, including pathogens, but never in environmental samples. For the remaining group of classic phytases detected in this study (PAP), the literature describes two branches, the PAP1 enzymes, which are Mg²⁺-dependent enzymes, and the PAP2 enzymes, which are Mg²⁺ independent, but in all cases the active forms of PAP phytases were derived from plants (25). We detected the PAPs Pho18 and Pho24, which are affiliated with bacteria and belong to the Mg²⁺-independent branch (PAP2 cl00474) (Fig. 1 and 2C).

Alpha/beta hydrolases (abhydrolases) represent a group of proteins with a high number of substrates and catalytic functions (26). Two gene products (Pho03B and Pho25C) contained an abhydrolase domain (Fig. 1). However, only Pho25C showed phosphatase/phytase activity after individual heterologous expression of the corresponding gene. Abhydrolases exhibit broad substrate specificity, and some members have been reported with phospholipase activity (27).

Other ORFs such as Pho16A carry the EAL domain, which is present in diverse bacterial signaling proteins and encodes a phosphodiesterase function (28). Analysis of Pho10 and Pho14D amino acid sequences indicates the presence of the P-loop_NTPase superfamily domain (Fig. 1). Enzymes harboring this domain hydrolyze the beta-gamma phosphate bond of, e.g., ATP and GTP (29). In this study, Pho10 showed phosphatase activity, while Pho14D as part of the clone harboring plasmid pLP14 showed none. Pho14C showed no phosphatase activity after individual heterologous expression of the corresponding gene. The *pho14C* gene product harbors the phospholipase D catalytic domain (PLDC_SF domain) (30).

SNARE-associated proteins with phosphatase activity harbor a new motif. In 19 out of 21 positive clones, we identified at least one gene encoding a protein domain

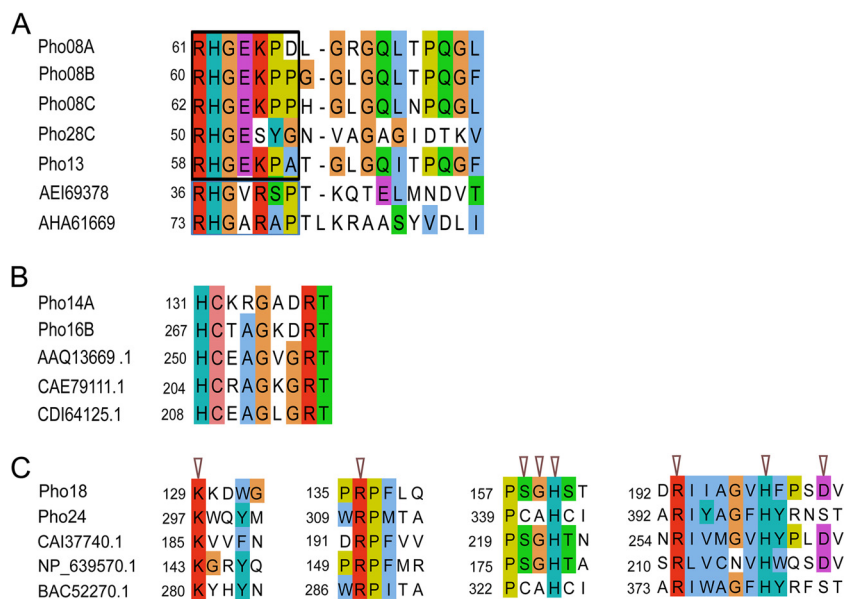


FIG 2 Multiple sequence alignments of conserved regions of phosphatases belonging to the HP, PTP, and PAP2 superfamily. (A) Blue line, typical conserved HP phytase motif (RHGXRXP) in AEI69378 (phytase from *Yersinia mollaretii*) and AHA61669 (histidine acid phosphatase phytase from *Thermothelomyces thermophila*). Black line, the variations of the motif found in this study. (B) Typical PTP motif (HCX5R) in Pho14A, Pho16B, AAQ13669 (myo-inositol hexaphosphate phosphohydrolase from *Selenomonas ruminantium*), CAE79111 (protein tyrosine phosphatase 2 from *Bdellovibrio bacteriovorus* HD100), and CDI64125 (protein tyrosine phosphatase from *Xylophilus ampelinus*). (C) Catalytic sites of the PAP2 superfamily (cl00474), in Pho18, Pho24, CAI37740 (putative phosphatase from *Corynebacterium jeikeium*), NP_639570 (phosphatase from *Xanthomonas campestris*), and BAC52270 (phosphatase from *Bradyrhizobium diazoefficiens*).

associated with catalytic activity of phosphatases. In contrast, the phosphatase-related genes of plasmids pLP04 and pLP15 did not encode known catalytic domains or signatures directly or indirectly associated with phosphatases. Clones carrying these plasmids showed significant phosphatase activity, and the products Pho04 and Pho15 showed sequence similarity to other previously reported proteins carrying the SNARE domain. However, both proteins shared overall sequence identity to previously reported phosphatases (Table 2). After individual heterologous expression of *pho04* and *pho15*, phosphatase activity was confirmed for both gene products. Pho04 and Pho15 hold the SNARE-associated domain DedA. SNARE-associated proteins are classified as structural proteins that function as a protein-protein interaction module (31). To our knowledge, no proteins with SNARE domains have been previously discovered to possess phosphatase activity.

We performed an alignment based on the *pho04* and *pho15* gene products, which revealed a shared conserved region (Fig. 3). Next, we analyzed all 56,539 sequences associated with the SNARE-associated Golgi proteins InterPro entry (IPR032816) with respect to motifs that were similar to those found in Pho04 and Pho15. A total of 905 sequences showed the conserved sequence pattern or a similar form. The sequence analysis revealed that Pho04 and Pho15 and the other 905 SNARE-associated (IPR032816) sequences share the particular amino acid arrangement ESSF(F/L/I/V)P. Notably, with respect to all analyzed proteins the identified motif was mostly from bacteria and detected outside the SNARE domain (cl00429) (examples are depicted in Fig. 3). Pho04 harbors the SNARE domain but shows 48% sequence identity to a putative membrane-associated alkaline phosphatase from *Acetobacter tropicalis*, while the closest phosphatase-related hit for Pho15 was an alkaline phosphatase from an *Acidobacteria* representative (43% identity) (Table 2).

ALP-like superfamily and non-plant-derived PAP representatives showing phytase activity. We selected the gene products of *pho07* and *pho18* for comprehensive

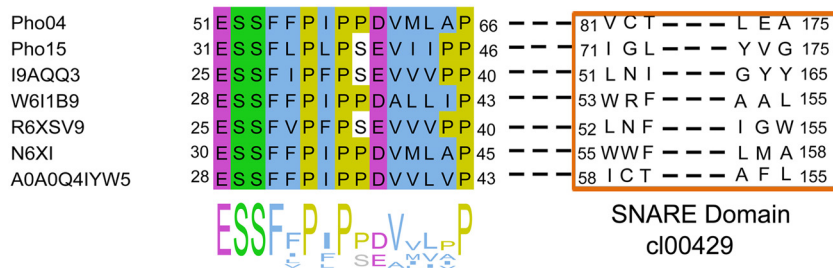


FIG 3 Partial multiple sequence alignment of Pho04, Pho15, and UniProt entries of SNARE-associated Golgi proteins. A detected conserved motif and its position in relation to the SNARE family are shown. The calculated consensus is depicted at the bottom. I9AQQ3, *Bacteroides fragilis*; W6I1B9, *Granulibacter thetensis*; R6XSV9, *Prevotella* sp.; N6XI35, *Thauera* sp.; and A0A0Q4IYW5, *Sphingomonas* sp.

biochemical characterization. The gene product of *pho07* does not contain any of the currently known catalytic domains associated with phytase activity. The only detected match of Pho07 was a nonspecific hit for the ALP-like superfamily (cl23718). In the case of *pho18*, the corresponding gene product comprises a domain of the purple acid phosphatases (PAP-like), which represents a type of phytase reported to be present in many organisms but is significantly expressed only in a very limited number of plant species (17, 32).

We successfully detected phytase activity of both purified enzymes, Pho07 and Pho18. Thus, to our knowledge Pho07 represents a new type of phytase and Pho18 represents the first PAP2 bacterial phytase. Furthermore, these two enzymes represent two out of the three reported environmental phytases derived from functional metagenomics. Both enzymes are putatively secreted by the natural bacterial host (Table S1) as the protein sequences harbor potential signal peptides of 30 (Pho07) and 22 (Pho18) amino acids at the N terminus. Pho07 shows the presence of an ALP-like superfamily domain (cl23718) (Fig. 1) and highest similarity to a phosphoesterase from a *Pseudonocardiales* representative (51% identity) (Table 2). Pho18 was most similar (50% identity) to an acid phosphatase from the *Verrucomicrobiaceae* member GAS474 (Table 2).

Pho07 and Pho18 exhibited optimal activity at 30 and 50°C, respectively (Fig. 4). After incubation of Pho07 for 4 h at 30°C, the enzyme retained more than 80% activity (Fig. S4). Incubation for 3 h at 45 and 60°C resulted in a substantial reduction (approximately 50%) and complete loss of enzyme activity, respectively. Pho18 retained

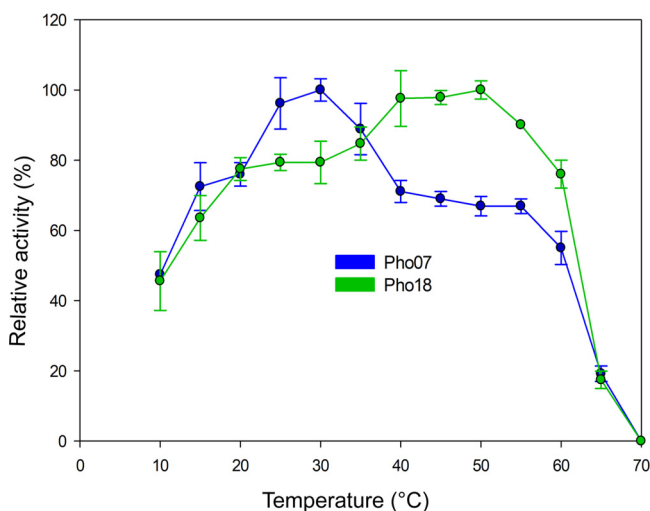


FIG 4 Effect of temperature on the relative activity of Pho07 and Pho18. All measurements were performed following the phytase standard assay at temperatures between 10 and 70°C. A 100% relative activity represented 2.9 and 1.04 U/mg for Pho07 and Pho18, respectively.

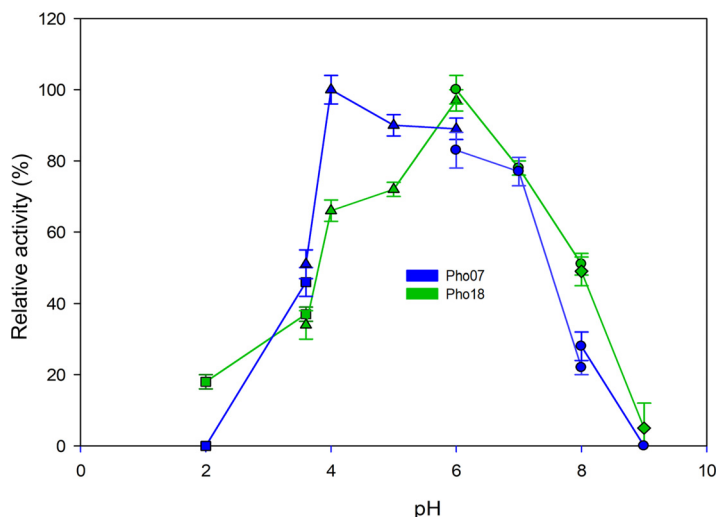


FIG 5 Effect of pH on the relative activity of Pho7 and Pho18. The measurements were performed with different buffer systems according to the phytase standard assay at the optimal temperature of each protein. The average from triplicate experiments is presented. Glycine-HCl buffer, squares; sodium acetate buffer, triangles; Tris-maleate buffer, circles; glycine-NaOH buffer, diamonds. 100% relative phytase activity represented 4.84 and 1.39 U/mg for Pho7 and Pho18, respectively.

approximately 80% activity after incubation for 6 h at 40°C but lost more than 50% of its activity at temperatures $\geq 50^\circ\text{C}$ (Fig. S4).

We evaluated the optimal pH range using different buffer systems at 30°C for Pho7 and at 50°C for Pho18. Pho7 exhibited the highest activity at pH 4.0 (Fig. 5) and retained more than 80% of its activity between pH 5.0 and 7.0. Low or no enzymatic activity was detected at pH values lower than 2.0 and higher than 8.0. Pho18 showed the highest activity at pH 6.0 and retained more than 70% of its activity at pH 5.0 and 7.0 (Fig. 5). To determine the substrate specificity of Pho7 and Pho18, we tested several phosphorylated compounds as the substrates (Fig. 6). Pho7 released phosphate from all tested compounds with the highest activity toward phytate and lowest activity toward pyrophosphate. Pho18 showed the highest relative activity with pyro-

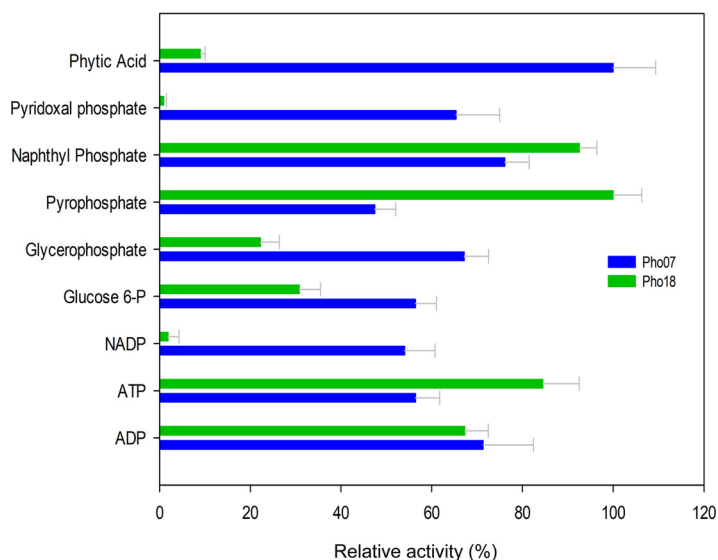


FIG 6 Substrate specificity of Pho7 and Pho18. Specific activities corresponding to 100% relative phytase and pyrophosphatase activities of Pho7 and Pho18 were 2.98 and 13.3 U/mg, respectively. All measurements were performed in triplicate and under optimal pH and temperature conditions for each enzyme.

TABLE 3 Kinetic values of Pho07 and Pho18 under optimal pH and temperature conditions

Enzyme	Mean (3 expts) ± SD							
	K_m (mM)		V_{max} ($\mu\text{mol min}^{-1} \text{mg}^{-1}$)		k_{cat} (min^{-1})		k_{cat}/K_m ($\text{min}^{-1} \text{M}^{-1}$)	
	Sodium phytate	Pyrophosphate	Sodium phytate	Pyrophosphate	Sodium phytate	Pyrophosphate	Sodium phytate	Pyrophosphate
Pho07	0.49 ± 0.18	1.09 ± 0.03	6.50E-03 ± 1.01E-06	1.30E-04 ± 8.05E-06	694 ± 12.43	516 ± 22.98	3,410 ± 122	4,991 ± 155
Pho18	0.96 ± 0.09	0.22 ± 0.04	2.82E-03 ± 2.01E-04	4.03E-04 ± 4.42E-07	152 ± 9.83	1,088 ± 34.09	1,550 ± 18	49,200 ± 274

phosphate as the substrate and no significant activity with pyridoxal phosphate and NADP. As Pho07 and Pho18 exhibited the highest activity with phytate and pyrophosphate, respectively, we used these substrates for calculation of kinetic constants (Table 3).

Finally, we measured the effect of various metal ions and potential enzyme inhibitors on the activity of Pho07 and Pho18 with phytate as the substrate (Fig. 7). The metal ions showed different effects on the activity of the analyzed proteins. Al^{3+} , Mn^{2+} , and Zn^{2+} increased the activity of Pho07, while the activity of Pho18 decreased in the presence of Zn^{2+} . Fe^{2+} had a strong inhibitory effect on the activity of both enzymes. With respect to potential inhibitors, the strongest inhibitory effects were observed at concentrations of 1 mM. Pho07 and Pho18 activities were reduced by most of the

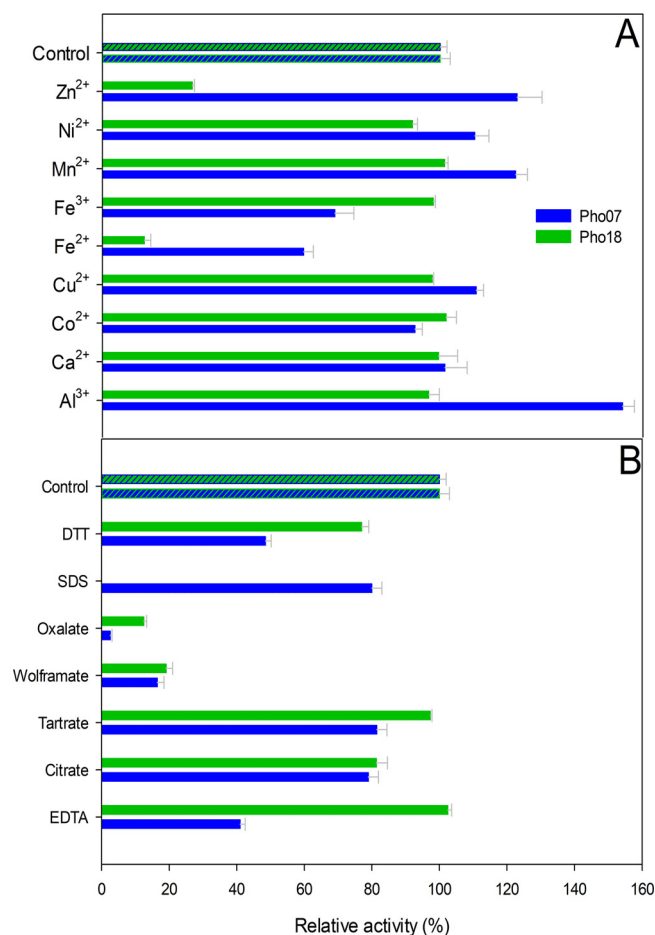


FIG 7 Effect of (A) metal ions and (B) potential inhibitors at 1 mM on the relative activity of Pho07 and Pho18. Specific activity values expressed as percentages of the control reactions are 3.8 and 1.3 U/mg for Pho07 and Pho18, respectively (A), and 3.5 and 1.22 U/mg for Pho07 and Pho18 (B), respectively.

tested inhibitors. Oxalate was the strongest inhibitor for Pho07, while the activity of Pho18 was completely depleted in the presence of SDS (Fig. 7).

DISCUSSION

Apart from the relationship with pathogenesis and the economic relevance, phosphatases/phytases are also important for reduction of phosphorus pollution and its impact on diverse environments (8, 11, 13). However, only a few phosphatases, most of them from cultivable organisms, have been comprehensively analyzed. The discovery of new phosphatases from environmental samples as well as engineering of available representatives of this enzyme group is considered a major research challenge (33). So far, few studies have attempted to discover phosphatases/phytases encoded by metagenomes using a function-based approach. Within these studies, only three genes and one of the corresponding proteins which exhibited phytase activity were recovered and described (34–36). We found 31 candidate genes, and 24 of them encoded phosphatase activity after individual heterologous expression (Table 2). For the remaining seven genes, activity was not detected at individual gene level. The corresponding gene products might be part of larger phosphatase units or require other components encoded by the insert to show phosphatase activity.

Approximately 55% of the gene products described in this study showed low protein sequence identity to known phosphatases (50% or less) (Table 2), which demonstrates the capacity of our screening method to identify novel enzymes with phosphatase activity from environmental samples. It has been previously discovered that the absence of free phosphate and the addition of phytate to medium induce the expression of phytases (37). Therefore, it is indicated that many of the detected genes encode new enzymes with phytase activity as observed for Pho07 and Pho18.

ALP phosphomonoesterases widely occur in nature. They preferably hydrolyze phosphate esters at pH levels higher than 7.0 (38). The ALP-like superfamily (cl23718) was the most abundant domain we detected in the recovered hits derived from our soil metagenomic libraries. The pH of the soil samples used ranged from 3.1 to 4.5 (39). Nevertheless, acid phosphatase genes are considered to be more abundant than alkaline phosphatase genes in low-pH soils. This might be due to the fact that most studies on the prevalence of alkaline and acid phosphatase genes are based on PCR-based gene amplification using specific known genes from cultured individual species as starting point for primer design (40). This approach covers only a small fraction of the existent functional phosphatase genes. Here, we revealed the existence of so-far-unknown functional ALPs with low identity toward known phosphatases, evidencing the potential of our functional metagenomic approach for the discovery of new ALP-phosphatases from environmental samples.

To our knowledge, enzymes from the ALP-like superfamily entry (cl23718) exhibiting phytase activity have not been described or comprehensively characterized yet. Nevertheless, numerous proteins are mentioned in literature or annotated in databases as alkaline phosphatases with phytase activity, but their molecular signatures and domains are associated mostly with the classic phytases (14). The analysis by Lim et al. (41) focusing on the distribution and diversity of phytate-mineralizing bacteria considers alkaline phosphatases to be ubiquitous in living organisms and shows that they dephosphorylate a wide range of P compounds, but not phytate. Thus, the functional proteins carrying the ALP-like superfamily domain reported in this study (7) represent a new group of phytase enzymes. The phylogenetic analysis of the ALP-like members revealed that most of our metagenome-derived enzymes cluster separately from previously reported alkaline phosphatases/phytases (see Fig. S2 in the supplemental material).

The biochemical analysis of a selected ALP-like member, Pho07, showed that its temperature optimum is similar to the metagenome-derived alkaline phosphatase (mAP). This enzyme is one of the few reported phosphatases derived from environmental samples and not associated with cultures (42). Furthermore, the optimal pH range of Pho07 (4.0 to 5.0) is similar to that of other soil bacterial phytases (43). Among

the tested substrates, Pho07 showed the highest activity toward phytate, indicating that its primary activity is related to the degradation of this compound. Several studies report an enhancing effect of Ca^{2+} and Mn^{2+} on phytase activity (43). Nevertheless, the activity of Pho07 increased in the presence of Mn^{2+} , but it was not affected by Ca^{2+} . Among the potential inhibitors, wolframate and oxalate did not show significant effects on the activity of a phytate-degrading enzyme from *Pantoea agglomerans* (44) but reduced the relative activity of Pho07 to values lower than 20%. Since Pho07 is the first reported phytase carrying an ALP-like domain, it is not possible to compare its kinetic parameters (Table 3) with those from phytases of the same type.

The enzyme Pho18 belongs to the known PAPphy group of phytases. Only a few examples of characterized PAP proteins with phytase activity have been previously reported, and all of them were derived from plants (25). However, the presence of PAP-related genes in mammals, fungi, and bacteria has been indicated based on annotated genome sequences. The taxonomic analysis of *pho18* and the complete insert harboring it revealed a bacterial origin and a phylogenetic association with the genus *Terrimicrobium* of the *Verrucomicrobia* phylum (Table S1). In addition, biochemical analysis confirmed phytase activity of Pho18. Therefore, we report here for the first time a PAP2 phosphatase with phytase activity, which is of nonplant origin and metagenome derived. Moreover, the phylogenetic analysis showed that Pho18 clusters separately from other previously reported PAPs with phytase activity. The reason for this is most likely the vegetal origin of the previously reported PAP phytases (Fig. S2). To our knowledge, the study of Ghorbani Nasrabadi et al. (45) is the only attempt to identify PAP phytases derived from bacteria. In their study, an indirect association between phytase activity and the amplification of a putative PAP gene in the bacterial host was established (45).

The optimal temperature of Pho18 (50°C) is similar to optimal temperatures of other PAPs derived from wheat (45°C) and soybeans (58°C) (14). Furthermore, the behavior of Pho18 at temperatures higher than 55°C (Fig. 4) is similar to that reported for soybean phytases (46). An increase of phytase activity mediated by the addition of Mn^{2+} was reported for PAP phytases (32, 43). We did not register significant increases in the activity of Pho18 in the presence of any cation. However, the enzyme was strongly inhibited by Zn^{2+} , which is in contrast to other PAP phytases showing higher activity in the presence of this ion. Although Pho18 exhibits higher affinity to pyrophosphate, the kinetic parameters using phytate as the substrate are similar to PAP phytases from *Arabidopsis* (Table 3) (47).

We found the HAD (cl21460) domain as the second most abundant domain in our survey. The HAD domain is present in proteins of diverse organisms, including bacteria, archaea, and eukaryotes (48). This domain is carried by proteins able to catalyze a variety of biological functions and act on a wide range of substrates (19). Numerous members of the HAD superfamily can transfer phosphoryl groups or act as phosphoanhydride hydrolase P-type ATPases (49). Since proteins harboring this domain are involved in a variety of cellular processes, it is not surprising that they can be isolated through functional metagenomic screening for phosphatases.

One of the most remarkable findings in this study was the detection of the SNARE-associated domain (DedA, InterPro entry IPR032816) of Pho04 and Pho15. So far, the role of the SNARE-associated domain (DedA) has not been deeply studied. Bacterial DedA family mutants display phenotypes evidencing cell division defects, temperature sensitivity, and altered membrane phospholipid composition among others (50). DedA-SNAREs have been reported to promote or block membrane fusion, particularly during bacterial pathogenic processes (51). To our knowledge no phosphatase activity has been reported for proteins harboring SNARE-associated domains. Moreover, the particular signature ESSF(F/L/I/V)P has been overlooked until now.

In conclusion, we demonstrate here for the first time the potential of functional metagenomics to exploit the phosphatase pools hidden in environmental samples. Our study revealed new phosphatases/phytases with diverse and, so far, largely unknown characteristics. Furthermore, we discovered the existence of a new type of phytases

(ALP-like-phy) and found that the classical PAPphy are also functional in microorganisms and not only in plants.

MATERIALS AND METHODS

Soil sampling, DNA extraction, and construction of metagenomic libraries. Genes encoding phosphatases were recovered from metagenomic libraries derived from A horizons of soil samples, which had been taken from forest sites of the German Biodiversity Exploratories Schwäbische Alb (samples AEW1 and AEW5), Hainich-Dün (sample HEW30), and Schorfheide-Chorin (samples SEW2, SEW5, and SEW46). Collection of samples was performed previously as described by Kaiser et al. (52) and Nacke et al. (39), respectively. Soil characteristics are available in Nacke et al. (39). Names of constructed metagenomic libraries refer to the designation of the samples from which the libraries were derived. Metagenomic libraries were generated using the method described by Nacke et al. (39). The plasmid libraries AEW1, AEW5, SEW2, and SEW5 have been previously generated by employing the same approach (39).

Function-based screening and identification of ORFs encoding phosphatase activity. For function-based screening of metagenomic libraries, we used our recently described method (18). Small-insert libraries were constructed using the plasmid pCR-XL-TOPO as vector (Invitrogen GmbH, Karlsruhe, Germany) and *Escherichia coli* DH5 α [F⁻ ϕ 80lacZ Δ M15 Δ (lacZYA-argF)U169 recA1 endA1 hsdR17(r_K⁻ m_K⁺) phoA supE44 λ -thi-1 gyrA96 relA1] as screening host. Modified Sperber medium (16 g/liter agar, 10 g/liter glucose, 500 mg/liter yeast extract, 100 mg/liter CaCl₂, and 250 mg/liter MgSO₄) was used as a screening medium supplemented with 2.5 g/liter phytic acid as sole P source and 25 μ g/ml of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) (53). The modified Sperber minimal medium used in this study was used for detection of phosphatase/phytase activity of the library-bearing *E. coli* clones due to the presence of phytate and the absence of other inorganic P sources (37). The slight background activity observed after more than 48 h of incubation of the host strain is probably caused by the alkaline phosphatase-encoding gene (*phoA*) of the host. Positive clones show an intense dark blue colony color, whereas negative colonies exhibit first a white and subsequently a light blue or green color after prolonged incubation.

The plasmids derived from positive clones were sequenced by the Göttingen Genomics Laboratory (Göttingen, Germany), and ORF prediction was performed as described by Nacke et al. (39). Next, the obtained sequences were analyzed by using the Basic Local Alignment Search Tool (BLAST) (54). Only plasmids harboring at least one ORF potentially associated with phosphatase activity were considered candidates for further analysis and designated pLP01 to pLP04, pLP07 to pLP10, pLP13 to pLP20, and pLP24 to pLP28. Full-length sequence alignment was performed between the candidates and their closest related sequence by using Clustal Omega (55). All coding sequences were examined for similarities to known protein families and domains by performing searches against the InterPro collection of protein signature databases and conserved domain databases (CDD) (56, 57). The prediction of signal peptides of the proteins was performed by using SignalP 4.0 (58). Additionally, all inserts were taxonomically classified by using the software Kaiju 1.5.0 (59). Alignments of the deduced protein sequences and phylogenetic trees of the proteins were performed by using MEGA 7 (60). The maximum likelihood method based on the equal input model was applied. The bootstrap values were calculated from 500 replicates, and branches corresponding to partitions reproduced in fewer than 50% of bootstrap replicates were collapsed. Alignments were visualized by using Jalview version 2 (61).

Candidate genes encoding domains that have not been previously associated with phosphatase activity (pLP04 and pLP15) and inserts comprising more than one potential phosphatase-encoding gene were amplified and subsequently cloned. Specific primers for each target gene were designed, and the pET101/D directional TOPO cloning kit was used for cloning as recommended by the manufacturer (Thermo Fisher Scientific GmbH, Schwerte, Germany). PCR was carried out in a 50- μ l volume containing 10 μ l of 5-fold Phusion GC buffer, 200 μ M (each) dNTP, 1.5 mM MgCl₂, 2 μ M (each) primers, 2.5% DMSO, 0.5 U Phusion High Fidelity Hot Start DNA polymerase (Thermo Fisher Scientific GmbH, Schwerte, Germany), and 25 ng recombinant plasmid. PCR conditions were as follows: initial denaturation at 98°C for 2 min followed by 30 cycles of denaturation at 98°C for 1.5 min, annealing at 58°C for 1 min, and extension at 72°C for 1 min, followed by a final extension at 72°C for 5 min. Subsequently, the amplified genes were individually cloned into the expression vector pET 101/D and transformed into *E. coli* BL21 [F⁻ *ompT hsdS_B*(r_B⁻ m_B⁻) *gal dcm* (DE3)] as recommended by the manufacturer (Thermo Fisher Scientific GmbH). The resulting recombinant plasmid-bearing *E. coli* BL21 strains were subsequently plated on Sperber minimal medium agar supplemented with phytic acid (2.5 g/liter), BCIP (25 mg/ml), and IPTG (0.25 mM) for phosphatase activity detection.

Heterologous expression of *pho07* and *pho18* and purification of gene products. The genes *pho07* and *pho18* carried by plasmids pLP07 and pLP18, respectively, were selected for heterologous expression with the pET-20b (+) (V5-epitope/His tag) vector (Merck KGaA, Darmstadt, Germany) as recommended by Villamizar et al. for metagenome-derived phosphatases (18). Crude extracts containing the target proteins were derived from the expression strain *E. coli* BL21 and filtered as described by Villamizar et al. (18). For purification of the proteins, the filtered crude extracts were then transferred to nickel columns (Protino2000 Ni-Ted columns; Macherey and Nagel, Düren, Germany). The equilibration of the columns and the washing steps were performed with 50 mM HEPES buffer (pH 8.0) containing 200 mM NaCl, followed by three elution steps with 50 mM HEPES, 200 mM NaCl, and 250 mM imidazole. Pho07 was further purified by using the Äkta FPLC system (GE Healthcare, Little Chalfont, United Kingdom) via hydrophobic interaction chromatography. A 15PHE 4.6/100PE Tricorn high-performance column (GE Healthcare, Little Chalfont, United Kingdom) in a total bed volume of 1.7 ml with a 2-ml/min

flow rate at room temperature was utilized. Pho18 was purified through ion-exchange chromatography, by using a cation exchanger (SOURCE15S) in a prepacked Tricorn column (4.6/100 PE) (GE Healthcare, Little Chalfont, United Kingdom) with a gel bed volume of 1.7 ml at a 1-ml/min flow rate and room temperature. The purity of the resulting protein preparations was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the detection of V5 epitope-carrying proteins was achieved by Western blot hybridization, as described by Waschkowitz et al. (62).

Enzyme assays. Phosphatase activity was determined at 355 nm by detecting the release of inorganic phosphorus according to the ammonium molybdate method developed by Heinonen and Lahti with modifications (44, 63) as follows: the enzyme solutions (10 μ l) were preincubated for 3 min at 40°C in 380 μ l of 50 mM sodium acetate buffer (pH 5). Subsequently, 10 μ l of 100 mM phytic acid dipotassium salt (Sigma-Aldrich, Munich, Germany) was added, and the mixture was incubated for 30 min at 40°C. To stop the reaction, 1.5 ml of freshly prepared AAM solution (acetone–5 N H₂SO₄–10 mM ammonium molybdate) and 100 μ l of 1 M citric acid were added. Blanks were prepared by adding AAM solution prior to the addition of enzyme. The absorbance (355 nm) was measured using the Ultrospec 3300 Pro (Amersham plc, Little Chalfont, United Kingdom).

To assess the influence of pH on purified enzymes, the activity was measured at 40°C in a pH range from 1 to 9. The following overlapping buffer systems were used: 50 mM glycine-HCl (pH 1.0 to 3.5), 50 mM sodium acetate (pH 3.5 to 6.0), 50 mM Tris-maleate acid (pH 6.0 to 8.0), and 50 mM glycine-NaOH (pH 7.0 to 9.0). After the optimal pH was determined for Pho07 and Pho18, the influence of temperature on enzymatic activity was analyzed. The thermal stability was checked after incubation of the purified enzymes at different temperatures.

The substrate specificity of the phosphatases was determined using the standard assay described above under the optimal temperature and pH for each enzyme (substrate concentration, 10 mM). Furthermore, the effects of cations (Al³⁺, Ca²⁺, Co²⁺, Fe²⁺, Fe³⁺, Mn²⁺, Ni²⁺, and Zn²⁺) and the potential inhibitors (EDTA, citrate, tartrate, wolframate, oxalate, sodium dodecyl sulfate (SDS), and dithiothreitol (DTT) at concentrations of 0.1 and 1 mM were analyzed.

For the kinetic constants, all measurements were performed in triplicate under optimal pH and temperature conditions using phytic acid and pyrophosphate as the substrates. The data were analyzed by the Sigma Plot Enzyme Kinetic Module version SigmaPlot 12.0 (Systat Software, Inc., San Jose, CA).

Sequence accession numbers. The nucleotide sequences of plasmids listed in Table 1 have been submitted to the National Center for Biotechnology Information (NCBI) GenBank database under the accession numbers indicated: pLP01 (Pho01), [KY931670](#); pLP02 (Pho02), [KY931671](#); pLP03 (Pho03A and -B), [KY931672](#); pLP04 (Pho04), [KY931673](#); pLP07 (Pho07), [KY931674](#); pLP08 (Pho08A to -C), [KY931675](#); pLP09 (Pho09C), [KY931676](#); pLP10 (Pho10), [KY931677](#); pLP13 (Pho13), [KY931678](#); pLP14 (Pho14A to -D), [KY931679](#); pLP15 (Pho15), [KY931680](#); pLP16 (Pho16A and -B), [KY931681](#); pLP17 (Pho17A), [KY931682](#); pLP18 (Pho18), [KY931683](#); pLP19 (Pho19A), [KY931684](#); pLP20 (Pho20B), [KY931685](#); pLP24 (Pho24), [KY931686](#); pLP25 (Pho25B and -C), [KY931687](#); pLP26 (Pho26), [KY931688](#); pLP27 (Pho27A and -B), [KY931689](#); and pLP28 (Pho28A and Pho28C), [KY931690](#).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.01966-18>.

FIG S1, PDF file, 0.5 MB.

FIG S2, PDF file, 1.1 MB.

FIG S3, PDF file, 0.1 MB.

FIG S4, PDF file, 0.2 MB.

TABLE S1, PDF file, 0.04 MB.

ACKNOWLEDGMENTS

We thank the Deutscher Akademischer Austauschdienst (DAAD) and Colciencias Colombia for the financial support to Genis Andrés Castillo in the frame of the fellowship program ALECOL. Additionally, we acknowledge support by the DFG and the Open Access Publication Funds of the Göttingen University.

We thank Silja Brady and Mechthild Bömeke for assistance with respect to expression and purification of the proteins, as well as Dominik Schneider for supporting the taxonomic classification of the inserts. We thank Elisabeth Gullans for her contribution to metagenomic library screening.

REFERENCES

- Daniel R. 2005. The metagenomics of soil. *Nat Rev Microbiol* 3:470–478. <https://doi.org/10.1038/nrmicro1160>.
- Nacke H, Daniel R. 2014. Approaches in metagenome research: progress and challenges, p 38–43. In Nelson KE (ed), *Encyclopedia of metagenomics*. Springer, New York, NY.
- Adam N, Perner M. 2018. Novel hydrogenases from deep-sea hydrothermal vent metagenomes identified by a recently developed activity-based screen. *ISME J* 12:1225–1236. <https://doi.org/10.1038/s41396-017-0040-6>.
- Berini F, Casciello C, Marcone GL, Marinelli F. 2017. Metagenomics: novel

- enzymes from non-culturable microbes. *FEMS Microbiol Lett* 364:fnx211. <https://doi.org/10.1093/femsle/fnx211>.
5. Blüher D, Laha D, Thieme S, Hofer A, Eschen-Lippold L, Masch A, Balcke G, Pavlovic I, Nagel O, Schonsky A, Hinkelmann R, Worner J, Parvin N, Greiner R, Weber S, Tissier A, Schutkowski M, Lee J, Jessen H, Schaaf G, Bonas U. 2017. A 1-phytase type III effector interferes with plant hormone signaling. *Nat Commun* 8:2159. <https://doi.org/10.1038/s41467-017-02195-8>.
 6. Chen MJ, Dixon JE, Manning G. 2017. Genomics and evolution of protein phosphatases. *Sci Signal* 10:eaag1796. <https://doi.org/10.1126/scisignal.aag1796>.
 7. Golovan SP, Meidinger RG, Ajakaiye A, Cottrill M, Wiederkehr MZ, Barney DJ, Plante C, Pollard JW, Fan MZ, Hayes MA, Laursen J, Hjorth JP, Hacker RR, Phillips JP, Forsberg CW. 2001. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat Biotechnol* 19:741–745. <https://doi.org/10.1038/90788>.
 8. Elser J, Bennett E. 2011. Phosphorus cycle: a broken biogeochemical cycle. *Nature* 478:29–31. <https://doi.org/10.1038/478029a>.
 9. Naylor RL, Hardy RW, Bureau DP, Chiu A, Elliott M, Farrell AP, Forster I, Gatlin DM, Goldburg RJ, Hua K, Nichols PD. 2009. Feeding aquaculture in an era of finite resources. *Proc Natl Acad Sci U S A* 106:15103–15110. <https://doi.org/10.1073/pnas.0905235106>.
 10. Scholz RW, Ulrich AE, Eilittä M, Roy A. 2013. Sustainable use of phosphorus: a finite resource. *Sci Total Environ* 461-462:799–803. <https://doi.org/10.1016/j.scitotenv.2013.05.043>.
 11. Zhang X, Li Z, Yang H, Liu D, Cai G, Li G, Mo J, Wang D, Zhong C, Wang H, Sun Y, Shi J, Zheng E, Meng F, Zhang M, He X, Zhou R, Zhang J, Huang M, Zhang R, Li N, Fan M, Yang J, Wu Z. 2018. Novel transgenic pigs with enhanced growth and reduced environmental impact. *Elife* 7:e34286. <https://doi.org/10.7554/eLife.34286>.
 12. Schröder JJ, Smit AL, Cordell D, Rosemarin A. 2011. Improved phosphorus use efficiency in agriculture: a key requirement for its sustainable use. *Chemosphere* 84:822–831. <https://doi.org/10.1016/j.chemosphere.2011.01.065>.
 13. Ushasree MV, Vidya J, Pandey A. 2017. Other enzymes: phytases, p 309–333. *In* Pandey A, Negi S, Soccol CR (ed), *Current developments in biotechnology and bioengineering*. Elsevier, Amsterdam, Netherlands.
 14. Lei XG, Weaver JD, Mullaney E, Ullah AH, Azain MJ. 2013. Phytase, a new life for an “old” enzyme. *Annu Rev Anim Biosci* 1:283–309. <https://doi.org/10.1146/annurev-animal-031412-103717>.
 15. Jiang X, Sobetzko P, Nasser W, Reverchon S, Muskhelishvili G. 2015. Chromosomal “stress-response” domains govern the spatiotemporal expression of the bacterial virulence program. *mBio* 6:e00353-15. <https://doi.org/10.1128/mBio.00353-15>.
 16. Tsang PW, Fong WP, Samaranyake LP. 2017. *Candida albicans* orf19.3727 encodes phytase activity and is essential for human tissue damage. *PLoS One* 12:e0189219. <https://doi.org/10.1371/journal.pone.0189219>.
 17. Mullaney EJ, Ullah AHJ. 2003. The term phytase comprises several different classes of enzymes. *Biochem Biophys Res Commun* 312: 179–184. <https://doi.org/10.1016/j.bbrc.2003.09.176>.
 18. Villamizar GAC, Nacke H, Daniel R. 2017. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity. *Methods Mol Biol* 1539:249–260. https://doi.org/10.1007/978-1-4939-6691-2_16.
 19. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, Toews Keating S, Aono M, Love JD, Evans B, Seidel RD, Hillerich BS, Garforth SJ, Almo SC, Mariano PS, Dunaway-Mariano D, Allen KN, Farelli JD. 2015. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc Natl Acad Sci U S A* 112:E1974–E1983. <https://doi.org/10.1073/pnas.1423570112>.
 20. Mustelin T. 2007. A brief introduction to the protein phosphatase families, p 9–22. *In* Moorhead G (ed), *Protein phosphatase protocols*. Springer, Totowa, NJ.
 21. Rigden DJ. 2008. The histidine phosphatase superfamily: structure and function. *Biochem J* 409:333–348. <https://doi.org/10.1042/BJ20071097>.
 22. Oh BC, Choi WC, Park S, Kim YO, Oh TK. 2004. Biochemical properties and substrate specificities of alkaline and histidine acid phytases. *Appl Microbiol Biotechnol* 63:362–372. <https://doi.org/10.1007/s00253-003-1345-0>.
 23. Puhl AA, Gruninger RJ, Greiner R, Janzen TW, Mosimann SC, Selinger LB. 2007. Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase. *Protein Sci* 16:1368–1378. <https://doi.org/10.1110/ps.062738307>.
 24. Gruninger RJ, Dobing S, Smith AD, Bruder LM, Selinger LB, Wieden HJ, Mosimann SC. 2012. Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases. *J Biol Chem* 287:9722–9730. <https://doi.org/10.1074/jbc.M111.309872>.
 25. Schenk G, Mitić N, Hanson GR, Comba P. 2013. Purple acid phosphatase: a journey into the function and mechanism of a colorful enzyme. *Coord Chem Rev* 257:473–482. <https://doi.org/10.1016/j.ccr.2012.03.020>.
 26. Nardini M, Dijkstra BW. 1999. α/β hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol* 9:732–737. [https://doi.org/10.1016/S0959-440X\(99\)00037-8](https://doi.org/10.1016/S0959-440X(99)00037-8).
 27. Holmquist M. 2000. Alpha beta-hydrolase fold enzymes structures, functions and mechanisms. *Curr Protein Pept Sci* 1:209–235. <https://doi.org/10.2174/1389203003381405>.
 28. Simm R, Morr M, Kader A, Nimitz M, Römmling U. 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol Microbiol* 53:1123–1134. <https://doi.org/10.1111/j.1365-2958.2004.04206.x>.
 29. Leipe DD, Koonin EV, Aravind L. 2004. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol* 343:1–28. <https://doi.org/10.1016/j.jmb.2004.08.023>.
 30. Waite M. 1999. The PLD superfamily: insights into catalysis. *Biochim Biophys Acta* 1439:187–197. [https://doi.org/10.1016/S1388-1981\(99\)00094-3](https://doi.org/10.1016/S1388-1981(99)00094-3).
 31. Fratti RA, Wickner W. 2007. Distinct targeting and fusion functions of the PX and SNARE domains of yeast vacuolar Vam7p. *J Biol Chem* 282: 13133–13138. <https://doi.org/10.1074/jbc.M700584200>.
 32. Dionisio G, Madsen CK, Holm PB, Welinder KG, Jorgensen M, Stoger E, Arcalis E, Brinch-Pedersen H. 2011. Cloning and characterization of purple acid phosphatase phytases from wheat, barley, maize, and rice. *Plant Physiol* 156:1087–1100. <https://doi.org/10.1104/pp.110.164756>.
 33. Kumar A, Chanderman A, Makolomakwa M, Perumal K, Singh S. 2016. Microbial production of phytases for combating environmental phosphate pollution and other diverse applications. *Crit Rev Environ Sci Technol* 46:556–591. <https://doi.org/10.1080/10643389.2015.1131562>.
 34. Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson AD, O’Gara F. 2014. Identification of novel phytase genes from an agricultural soil-derived metagenome. *J Microbiol Biotechnol* 24:113–118. <https://doi.org/10.4014/jmb.1307.07007>.
 35. Lee MH, Oh KH, Kang CH, Kim JH, Oh TK, Ryu CM, Yoon JH. 2012. Novel metagenome-derived, cold-adapted alkaline phospholipase with superior lipase activity as an intermediate between phospholipase and lipase. *Appl Environ Microbiol* 78:4959–4966. <https://doi.org/10.1128/AEM.00260-12>.
 36. Farias N, Almeida I, Meneses C. 2018. New bacterial phytase through metagenomic prospection. *Molecules* 23:448. <https://doi.org/10.3390/molecules23020448>.
 37. Kerovuo J, Lauraeus M, Nurminen P, Kalkkinen N, Apajalahti J. 1998. Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Appl Environ Microbiol* 64:2079–2085.
 38. Sharma U, Pal D, Prasad R. 2014. Alkaline phosphatase: an overview. *Indian J Clin Biochem* 29:269–278. <https://doi.org/10.1007/s12291-013-0408-y>.
 39. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol Ecol* 78:188–201. <https://doi.org/10.1111/j.1574-6941.2011.01088.x>.
 40. Fraser T, Lynch DH, Entz MH, Dunfield KE. 2015. Linking alkaline phosphatase activity with bacterial phoD gene abundance in soil from a long-term management trial. *Geoderma* 257:115–122. <https://doi.org/10.1016/j.geoderma.2014.10.016>.
 41. Lim BL, Yeung P, Cheng C, Hill JE. 2007. Distribution and diversity of phytate-mineralizing bacteria. *ISME J* 1:321–330. <https://doi.org/10.1038/ismej.2007.40>.
 42. Lee DH, Choi SL, Rha E, Kim SJ, Yeom SJ, Moon JH, Lee SG. 2015. A novel psychrophilic alkaline phosphatase from the metagenome of tidal flat sediments. *BMC Biotechnol* 15:1. <https://doi.org/10.1186/s12896-015-0115-2>.
 43. Suleimanova AD, Beinbauer A, Valeeva LR, Chastukhina IB, Balaban NP, Shakirov EV, Greiner R, Sharipova MR. 2015. Novel glucose-1-phosphatase with high phytase activity and unusual metal ion activation from soil

- bacterium *Pantoea* sp. strain 3.5.1. *Appl Environ Microbiol* 81: 6790–6799. <https://doi.org/10.1128/AEM.01384-15>.
44. Greiner R. 2004. Purification and properties of a phytate-degrading enzyme from *Pantoea agglomerans*. *Protein J* 23:567–576. <https://doi.org/10.1007/s10930-004-7883-1>.
 45. Ghorbani Nasrabadi R, Greiner R, Yamchi A, Nourzadeh Roshan E. 2018. A novel purple acid phytase from an earthworm cast bacterium. *J Sci Food Agric* 98:3667–3674. <https://doi.org/10.1002/jsfa.8845>.
 46. Hegeman CE, Grabau EA. 2001. A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *Plant Physiol* 126:1598–1608. <https://doi.org/10.1104/pp.126.4.1598>.
 47. Kuang R, Chan KH, Yeung E, Lim BL. 2009. Molecular and biochemical characterization of AtPAP15, a purple acid phosphatase with phytase activity, in *Arabidopsis*. *Plant Physiol* 151:199–209. <https://doi.org/10.1104/pp.109.143180>.
 48. Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. 2006. Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J Mol Biol* 361:1003–1034. <https://doi.org/10.1016/j.jmb.2006.06.049>.
 49. Koonin EV, Tatusov RL. 1994. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. *J Mol Biol* 244:125–132. <https://doi.org/10.1006/jmbi.1994.1711>.
 50. Doerrler WT, Sikdar R, Kumar S, Boughner LA. 2013. New functions for the ancient DedA membrane protein family. *J Bacteriol* 195:3–11. <https://doi.org/10.1128/JB.01006-12>.
 51. Wesolowski J, Paumet F. 2010. SNARE motif: a common motif used by pathogens to manipulate membrane fusion. *Virulence* 1:319–324. <https://doi.org/10.4161/viru.1.4.12195>.
 52. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrupf M, Daniel R. 2016. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci Rep* 6:33696. <https://doi.org/10.1038/srep33696>.
 53. Sperber JI. 1958. The incidence of apatite-solubilizing organisms in the rhizosphere and soil. *Aust J Agric Res* 9:778–781. <https://doi.org/10.1071/AR9580778>.
 54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 55. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
 56. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
 57. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH. 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 41:D348–D352. <https://doi.org/10.1093/nar/gks1243>.
 58. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>.
 59. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>.
 60. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
 61. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
 62. Waschowitz T, Rockstroh S, Daniel R. 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microbiol* 75: 2506–2516. <https://doi.org/10.1128/AEM.02136-08>.
 63. Heinonen JK, Lahti RJ. 1981. A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase. *Anal Biochem* 113:313–317. [https://doi.org/10.1016/0003-2697\(81\)90082-8](https://doi.org/10.1016/0003-2697(81)90082-8).

Supplemental Information for Chapter 3

Contents

Fig. S1. Sperber medium indicator plate containing the negative control (*E. coli* DH5 α carrying pCR-XL-TOPO) and a typical positive *E. coli* clone (*E. coli* DH5 α carrying plasmid pLP03).

Fig. S2. Phylogenetic trees of the retrieved phosphatases. Colors of each phylogenetic tree are consistent with the designated colors to the domains in Fig. 1.

Fig. S3. Insert of the plasmid pLP08 showing the tandem organization and relative position of the candidate genes pho08 A, B and C.

Fig. S4. Thermal stability of Pho07 (a) and Pho18 (b). All measurements were performed following the phytase standard assay. Specific activities corresponding to 100% relative phytase activity are 3.14 (a) and 1.61 (b) U/mg. The average of triplicate experiments is presented.

Table S1. Taxonomic classification of inserts from the positive clones harboring phosphatase-related genes by using KAIJU 1.5.0.

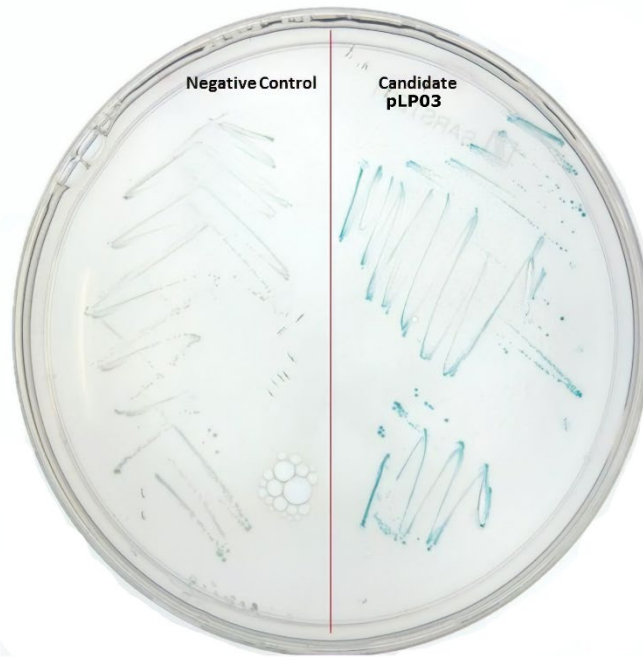


Fig. S1. Sperber medium indicator plate containing the negative control (*E. coli* DH5 α carrying pCR-XL-TOPO) and a typical positive *E. coli* clone (*E. coli* DH5 α carrying plasmid pLP03).

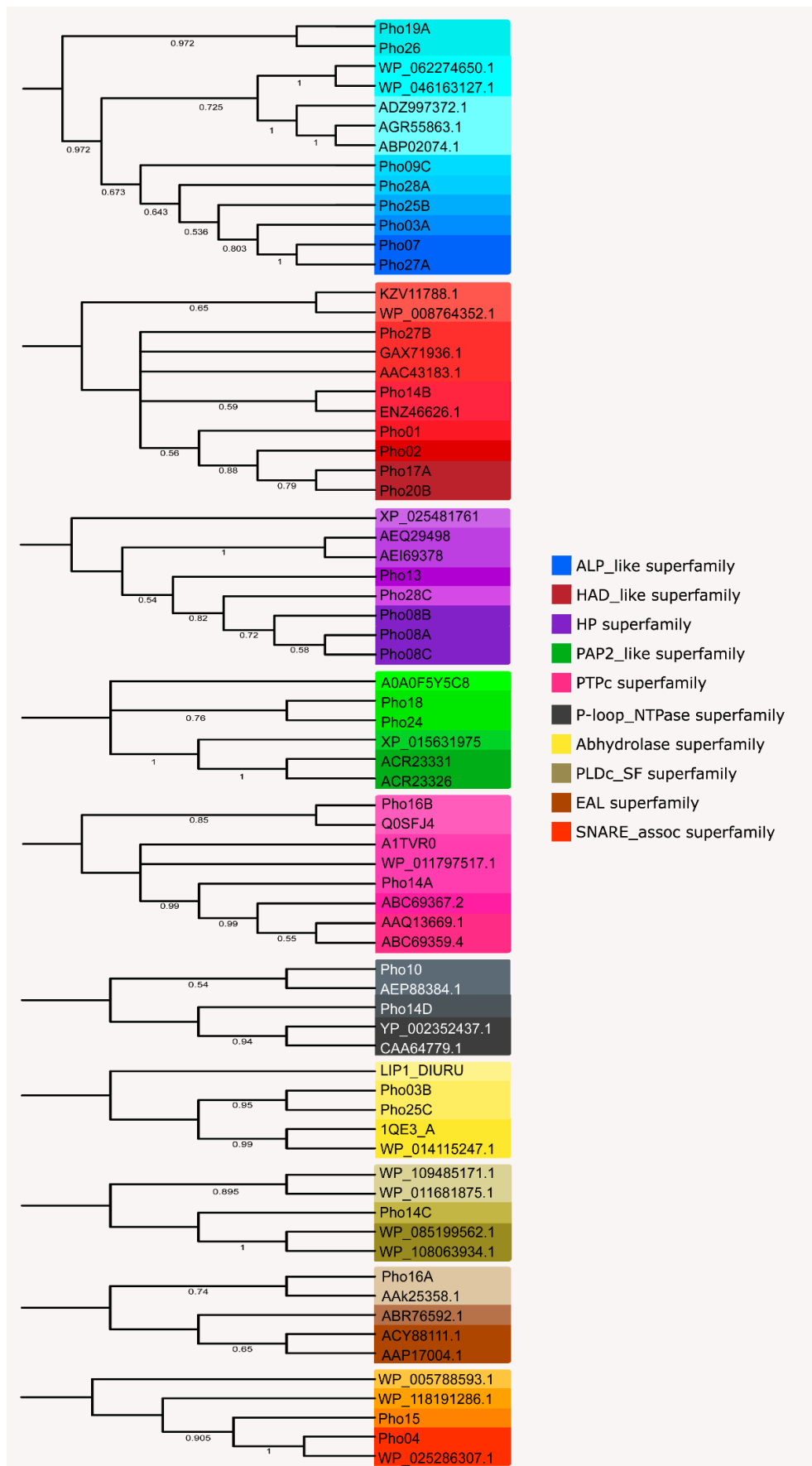


Fig. S2. Phylogenetic trees of the retrieved phosphatases. Colors of each phylogenetic tree are consistent with the designated colors to the domains in Fig. 1.

ALP-like (Alkaline phosphatases and sulfatases cl23718): AGR55863.1 alkaline phytase *Bacillus subtilis*, ADZ99372.1 beta-propellar phytase *Bacillus* sp, ABP02074.1 3-phytase *Bacillus licheniformis*, WP_046163127.1 alkaline phosphatase *Pseudomonas*, WP_062274650.1 alkaline phosphatase *Rhizobium*; HAD: (Haloacid dehalogenase superfamily cl21460): KZV11788.1 enzyme PHM8 *Saccharomyces cerevisiae*, WP_008764352.1 Cof-type HAD-IIB hydrolase *Bacteroides*, GAX71936 glycerol-1-phosphatase *Saccharomyces cerevisiae*, AAC43183.1 E-1 enzyme from *Klebsiella oxytoca*, ENZ46626 phosphatase *Clostridium bolteae*; HP (Histidine phosphatase cl11399): AEQ29498.1 histidine acid phytase, partial *Serratia* sp, AEI69378.1 phytase *Yersinia mollaretii*, XP_025481761.1 3-phytase *Aspergillus niger*; PAP2 (Phosphatidic acid phosphatase cl00474): XP_015631975.1 purple acid phosphatase *Oryza sativa*, ACR23331.1 purple acid phosphatase *Hordeum vulgare*, ACR23326.1 purple acid phosphatase *Triticum aestivum*; PTPs (Protein tyrosine phosphatases cl21483): Q0SFJ4 Possible tyrosine protein phosphatase *Rhodococcus jostii*, WP_011797517.1 PTP *Acidovorax citrulli*, P96830 PTP *Mycobacterium tuberculosis*, ABC69367 protein tyrosine phosphatase-like inositol polyphosphate phosphatase *Selenomonas lactificex*, AAQ13669.1 myo-inositol hexaphosphate phosphohydrolase *Selenomonas ruminantium*, ABC69359.4 PTP-like phytase *Selenomonas ruminantium*; P-loop_NTPase superfamily cl21455: AEP88384.1 protein YvcJ *Bacillus subtilis*, Mrp protein *Dictyoglomus turgidum*, CAA64779.1 Nbp35p protein *Saccharomyces cerevisiae*; Abhydrolase superfamily cl21494: LIP1_DIURU *Diutina rugosa*, 1QE3_A para-nitrobenzyl esterase *Bacillus subtilis*, WP_014115247.1 carboxylesterase/lipase family protein *Bacillus*; PLDC: Phospholipase D cl15239: WP_109485171.1 phosphatidylserine synthase *Occallatibacter savannae*, WP_011681875.1 phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase-like *Candidatus Solibacter usitatus*, WP_085199562.1 phospholipase *Mycobacterium fragae*, WP_108063934.1 cardiolipin synthase *Spartobacteria*; EAL superfamily cl00290: ACY88111.1 protein STM14_1632 *Salmonella enterica*, ABR76592.1 hypothetical protein KPN_01159 *Klebsiella pneumoniae*, AAP17004.1 hypothetical protein S1641 *Shigella flexneri*, AAK25358.1 GGDEF family protein *Caulobacter vibrioides*. SNARE-associated superfamily cl00429: WP_005788593.1 DedA protein *Bacteroides*, WP_025286307.1 DedA protein *Granulibacter bethesdensis*, WP_118191286.1 DedA protein *Prevotella copri*.

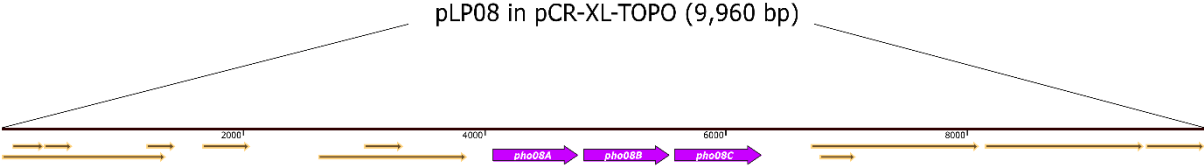


Fig. S3. Insert of the plasmid pLP08 showing the tandem organization and relative position of the candidate genes *pho08 A, B and C*.

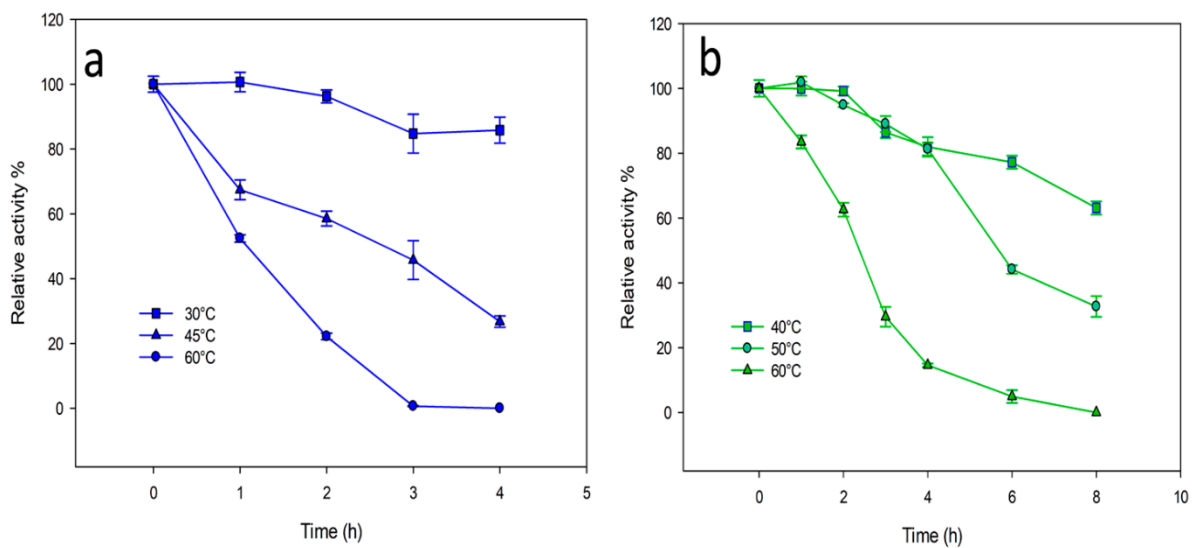


Fig. S4. Thermal stability of Pho07 (a) and Pho18 (b). All measurements were performed following the phytase standard assay. Specific activities corresponding to 100% relative phytase activity are 3.14 (a) and 1.61 (b) U/mg. The average of triplicate experiments is presented.

Table S1. Taxonomic classification of inserts from the positive clones harboring phosphatase-related genes by using KAIJU 1.5.0.

Plasmid	Taxonomic classification of insert
pLP01	Cellular organisms; Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Rhodanobacteraceae; Rhodanobacter; <i>Rhodanobacter</i> sp. C03
pLP02	Cellular organisms; Bacteria; environmental samples; uncultured bacterium
pLP03	Cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Actinobacteria; Streptomyetales; Streptomyetaceae; Streptacidiphilus; <i>Streptacidiphilus jiangxiensis</i>
pLP04	Cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; unclassified Caulobacterales; Caulobacterales <i>bacterium</i> RIFOXYB1_FULL_67_16
pLP07	Cellular organisms; Bacteria; Terrabacteria group
pLP08	Cellular organisms; Bacteria; Proteobacteria; Oligoflexia; Bdellovibrionales
pLP09	Cellular organisms; Bacteria; PVC group; Verrucomicrobia; unclassified Verrucomicrobia; unclassified Verrucomicrobia (miscellaneous); Verrucomicrobia <i>bacterium</i> RIFCSPLOWO2_12_FULL_64_8
pLP10	Cellular organisms; Bacteria
pLP13	Cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Actinobacteria; Corynebacteriales; Mycobacteriaceae; Mycobacterium; environmental samples; uncultured <i>Mycobacterium</i> sp.
pLP14	Cellular organisms; Bacteria; Terrabacteria group; Chloroflexi; unclassified Chloroflexi; unclassified Chloroflexi (miscellaneous); <i>Chloroflexi bacterium</i> 13_1_40CM_55_7
pLP15	Cellular organisms; Bacteria; PVC group; Verrucomicrobia; Opitutae; unclassified Opitutae; Opitutae <i>bacterium</i> TMED102
pLP16	Cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Actinobacteria; Streptomyetales; Streptomyetaceae; Streptacidiphilus; <i>Streptacidiphilus melanogenes</i>
pLP17	Cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; <i>Bradyrhizobium</i>
pLP18	Cellular organisms; Bacteria; PVC group; Verrucomicrobia; Spartobacteria; Terrimicrobium; <i>Terrimicrobium sacchariphilum</i> ;
pLP19	Cellular organisms; Bacteria
pLP20	Cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium; <i>Bradyrhizobium paxllaeri</i>
pLP24	Cellular organisms; Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium; <i>Bradyrhizobium</i> sp. AS23.2
pLP25	Cellular organisms; Bacteria; PVC group; Verrucomicrobia; unclassified Verrucomicrobia; unclassified Verrucomicrobia (miscellaneous)
pLP26	Cellular organisms; Bacteria; Acidobacteria; Acidobacteriia; Acidobacteriales; Acidobacteriaceae; unclassified Acidobacteriaceae; Acidobacteriaceae <i>bacterium</i> KBS 89
pLP27	Cellular organisms; Bacteria; Terrabacteria group; Chloroflexi; Ktedonobacteria; Ktedonobacterales; Ktedonobacteraceae; <i>Ktedonobacter</i>
pLP28	Cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Actinobacteria; Streptomyetales; Streptomyetaceae; Streptomyces; <i>Streptomyces griseoplanus</i>

4**Characteristics of the First Protein Tyrosine Phosphatase with Phytase Activity from a Soil Metagenome**

Genis Andrés Castillo Villamizar^{1,2}, Heiko Nacke¹, Laura Griese¹, Lydia Tabernero³,
Katrina Funkner¹ and Rolf Daniel¹

Genes (2019), 10:101

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany

²Línea tecnológica biocorrosión, Corporación para la investigación de la corrosión C.I.C. Piedecuesta, Santander, Colombia

³School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester

Author contributions to the work:

Performed the experiments: GAC, LG, KF.

Analyzed data: GAC, HN, LT.

Wrote the paper: GAC, HN, LT, RD

Conceived and designed the experiments: HN, RD.

Article

Characteristics of the First Protein Tyrosine Phosphatase with Phytase Activity from a Soil Metagenome

Genis Andrés Castillo Villamizar ^{1,2}, Heiko Nacke ¹, Laura Griese ¹, Lydia Taberero ³, Katrina Funkner ¹ and Rolf Daniel ^{1,*}

¹ Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University of Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany; gcastil@gwdg.de (G.A.C.V.); hnacke@gwdg.de (H.N.); lauragriese@live.de (L.G.); katrina.funkner@stud.uni-goettingen.de (K.F.)

² Línea Tecnológica Biocorrosión, Corporación para la investigación de la corrosión C.I.C. Piedecuesta, Santander, Colombia

³ School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; lydia.taberero@manchester.ac.uk

* Correspondence: rdaniel@gwdg.de; Tel.: +49-551-39-33827

Received: 21 December 2018; Accepted: 24 January 2019; Published: 29 January 2019

Abstract: Protein tyrosine phosphatases (PTPs) fulfil multiple key regulatory functions. Within the group of PTPs, the atypical lipid phosphatases (ALPs) are known for their role as virulence factors associated with human pathogens. Another group of PTPs, which is capable of using inositol-hexakisphosphate (InsP₆) as substrate, are known as phytases. Phytases play major roles in the environmental phosphorus cycle, biotechnology, and pathogenesis. So far, all functionally characterized PTPs, including ALPs and PTP-phytases, have been derived exclusively from isolated microorganisms. In this study, screening of a soil-derived metagenomic library resulted in identification of a gene (*pho16B*), encoding a PTP, which shares structural characteristics with the ALPs. In addition, the characterization of the gene product (Pho16B) revealed the capability of the protein to use InsP₆ as substrate, and the potential of soil as a source of phytases with so far unknown characteristics. Thus, Pho16B represents the first functional environmentally derived PTP-phytase. The enzyme has a molecular mass of 38 kDa. The enzyme is promiscuous, showing highest activity and affinity toward naphthyl phosphate (K_m 0.966 mM). Pho16B contains the HCXXGKDR[TA]G submotif of PTP-ALPs, and it is structurally related to PtpB of *Mycobacterium tuberculosis*. This study demonstrates the presence and functionality of an environmental gene codifying a PTP-phytase homologous to enzymes closely associated to bacterial pathogenicity.

Keywords: metagenomics; phosphatases; phytases; promiscuous enzymes; metagenomic library

1. Introduction

Since the emergence of next generation sequencing and omics approaches, the genetic material of numerous organisms, including bacterial, plant, and animal pathogens, has been analyzed. Genome analysis of single organisms, together with metagenomic and metaproteomic surveys comprising diverse environmental samples, provided an improved understanding of microbial biodiversity and the relationship of diversity with ecological, biotechnological, evolutionary or pathogenic processes [1–3].

Protein tyrosine phosphatases (PTPs) are an example of enzymes associated with virulence whose environmental homologues have not fully been studied. PTPs have multiple roles in cell metabolism. For many PTPs, their physiological substrate has not been identified [4]. A major role of the PTPs includes regulating, together with the protein kinases, the cellular equilibrium of protein tyrosine phosphorylation by dephosphorylating tyrosine residues of protein substrates. PTPs also participate in cell signaling by dephosphorylating proteins on other amino acidic residues (serine and threonine), as well as lipid substrates [4,5]. Another group of PTP proteins, known as atypical lipid phosphatases (ALPs), are associated with different levels of metabolic control of phosphoinositides (PIs) [6]. ALPs possess a characteristic catalytic profile, sequence, and domain organization. They harbor a distinct active site P-loop signature (HCXXGKDR[TA]G), containing the acid/base catalyst (D) and an extra basic residue (K) important in substrate binding [6,7]. PTP-ALPs with this P-loop are found in bacteria, including human pathogens [6], but only very few have been characterized.

Other members of the PTPs can hydrolyze *myo*-inositol phosphates (InsPs), which are ubiquitous products of inositol metabolism and bear a high level of structural resemblance to PIs [8]. The phytic acids (*myo*-inositol hexakisphosphate, InsP₆) have several roles within eukaryotic cells, including second messengers and cofactors that facilitate the regulation of diverse biochemical processes, such as transcription and hormone receptor activity [9,10]. A variety of other important biological functions have been directly or indirectly related to the presence of InsP₆. It has been reported that InsP₆ acts as a signal in the maintenance of basal resistance to viruses and phytopathogens [11]. Other reported functions of InsP₆ include antioxidative functions or involvement in DNA repair in prostate cancer prevention [12,13]. InsPs are typically absent in prokaryotes, but several genes encoding phosphatases (phytases) capable of using InsP₆ as substrate have been detected in bacteria [14–16].

InsP₆ also represents the main phosphorus (P) storage in many types of plants. It is considered a significant part of organic soil phosphate (Po) and relevant for the P cycle in soils [17]. Phytases decompose InsP₆ to less phosphorylated *myo*-inositol derivatives and inorganic phosphate [18]. Phytases are used as an effective feed additive that increases the digestion/absorption rates of phosphorus in cereal-based feed. In this way, livestock growth increases and phosphorus pollution caused by the non-assimilated InsP₆ are reduced [19]. The annual phytase market value is roughly \$700 million. Therefore, the search for new phytases has become a major research challenge [20]. However, few of the phytases—all derived from cultured microorganisms—have been commercially exploited.

Currently, four classes of phytases have been described: histidine acid phytase (HAPs-phy), β -propeller phytase (BPPs-phy), purple acid phytase (PAPs-phy), and protein tyrosine phytase (PTPs-phy). These enzymes are not structurally similar and use different mechanisms to cleave phosphate groups from InsP₆ [18,21,22]. PTPs have been relatively well studied, but not many enzymes of this type with phytase activity have been reported. Moreover, most of these phytases are derived from culturable anaerobic bacteria [21,23].

Several functional gene homologs of physiologically relevant kinases/phosphatases have been found as ubiquitous in non-human environments, demonstrating the relevant ecological roles of these genes for the activity and survival of environmental bacteria [3]. Although homologous sequences of PTPs, including ALPs, are present in the genome sequences of numerous microorganisms (cultured and uncultured), to our knowledge, characterized PTPs from environmental samples have not been reported. In addition, PTPs carrying the specific P-loop of the ALPs with phytase activity are also unknown. Here, we report, for the first time, a functional metagenome-derived PTP protein carrying the specific P-loop of ALPs and exhibiting phytase activity.

2. Results and Discussion

2.1. Identification and In Silico Characterization of the Novel ALP Member Pho16B

Most of the research on environmental phytases is represented by a few studies performing sequence-based identification of putative phytase genes in metagenomes. A limited number of the corresponding phytase proteins have been characterized. In addition, only three phytase-encoding genes derived from functional metagenomic approaches have been reported [24–27]. We have previously reported, to date, the highest number of environmentally derived phosphatases [28]. Function-based screening of a forest soil metagenomic library for genes coding for phosphatases and phytases yielded a positive *Escherichia coli* clone carrying the recombinant plasmid pLP16. The insert of this plasmid (7806 bp) harbors 40 open reading frames (ORFs). For 20 of the detected ORFs, it was possible to assign a putative function by similarity searches (Figure 1, Table S1).

The taxonomic classification of the complete pLP16 insert sequence is affiliated to the Gram-positive soil bacterium *Streptacidiphilus melanogenes*, originally isolated from *Pinus*-associated soils [29], and other Actinobacteria. Therefore, an actinobacterial relative of *S. melanogenes* might represent the original source of the cloned fragment. Consistently, the deduced protein sequence-based analysis showed that the gene product of *pho16B*, encoding a putative phytase, is closely related to the PTP WP_042381880 from *S. melanogenes* (79% of sequence identity). In addition to protein sequences stored in the National Center for Biotechnology Information (NCBI) nr database, the Pho16B sequence was compared with protein sequences of specialized metagenome databases. The closest related match to Pho16B in the latter databases was entry MGY000356208135 of the European Bioinformatics Institute (EMBL-EBI) metagenomics platform. MGY000356208135 shows 42% sequence identity to Pho16B. The corresponding gene sequence was derived from a sludge metagenome associated to an oil refinery and is similar to a gene encoding a PTP from *Sphingomonas* sp. (UniProt entry A0A1M2ZKN6).

The sequence analysis of *pho16B* revealed that it encodes a PTP, which contains the characteristic ALP signature (HCXXGKDR[TA]G). The presence of the conserved Lys and Asp residues indicates similarity of the ALPs and Pho16B with respect to the predicted catalytic mechanism [6,7] (Figure 2).

Pho16B contains a signal peptide with a predicted cleavage site between amino acid positions 36 and 37, indicating secretion of the protein and extracellular localization in the original host. Pho16B carries the PTPc domain and the specific Y_phosphatase3 domain (Figure 2A). The *pho16B* gene was subcloned in the expression vector pBAD202/D-TOPO and expressed in the *E. coli* host strain LMG194. Subsequently, the *pho16B* gene product was purified by a combination of affinity and hydrophobic interaction chromatography, yielding 155 µg pure enzyme from 500 mL culture, with a specific activity of 8.09 U/mg. The protein has a calculated molecular mass of approx. 38 kDa, which is similar to the molecular masses of other PTP-ALPs, such as lmo1800 and lmo1935 from *Listeria monocytogenes*, and phytases like Bd1204 of *Bdellovibrio bacteriovorus* [30].

We analyzed the phylogenetic position of Pho16B (Figure 3) by considering the phosphatase groups (G1–G9) established by Beresford et al. 2010 [6]. Bacterial sequences are clustered in the groups G1 to G6, while the eukaryotic sequences are clustered in the groups G7 to G9. The P96830 (MptpB protein of *Mycobacterium tuberculosis*) belongs to G1 together with other mycobacterial ALPs, whereas Pho16B belongs to G4, which comprises phosphatases of Gram-positive and Gram-negative bacteria. The closest related protein of Pho16B was included in our phylogenetic analysis. This protein, belonging also to G4, was annotated during genome analysis of *S. melanogenes* (accession WP_042381880). To our knowledge, the expression or characterization of the gene product have not been reported.

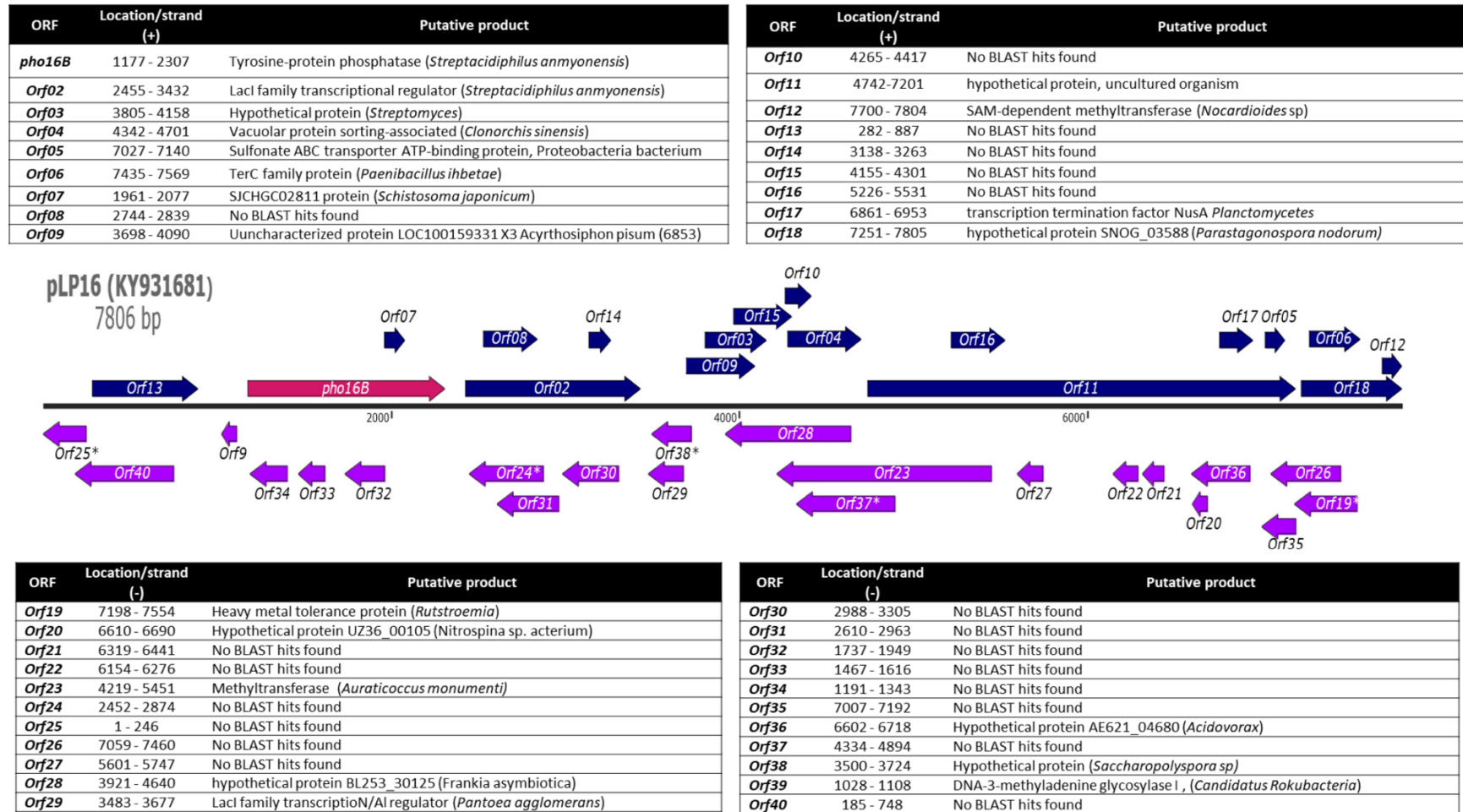


Figure 1. Strand, location, and BLAST results for all identified open reading frames (ORFs) of the pLP16 insert. * Partial ORFs.

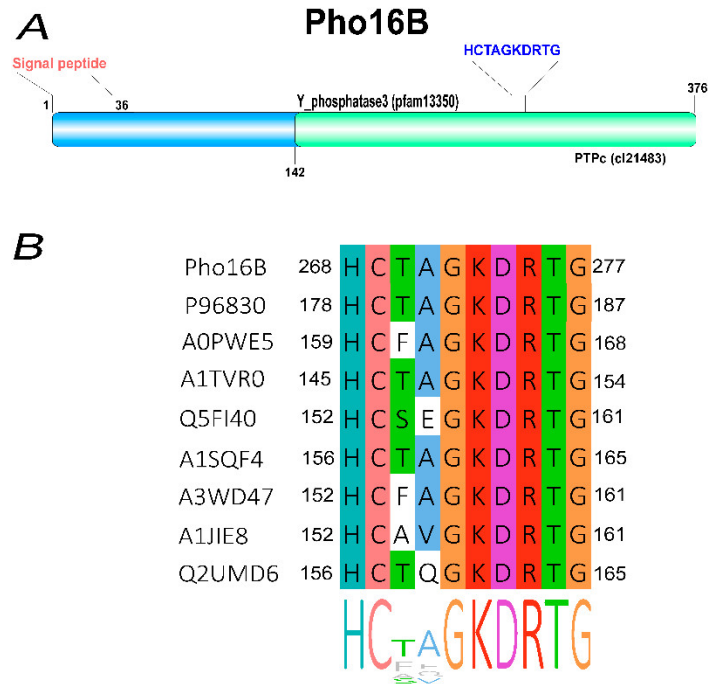


Figure 2. Domain organization and alignment of the subloop in Pho16B (A). Architecture of Pho16B showing the positions of the protein tyrosine phosphatase (PTP) domain and the signal peptide. (B) Alignment and position of the P-loop motif (HCXXGKDRTG) in Pho16B and other related atypical lipid phosphatase (ALP) proteins. Pho16B (this study), UniProtKB codes: P96830 (*Mycobacterium tuberculosis*), A0PWE5 (*Mycobacterium ulcerans*), A1TVR0 (*Acidovorax citrulli*), Q5FI40 (*Lactobacillus acidophilus*), A1SQF4 (*Nocardioides* sp.), A3WD47 (*Erythrobacter* sp.), A1JIE8 (*Yersinia enterocolitica*), and Q2UMD6 (*Aspergillus oryzae*).

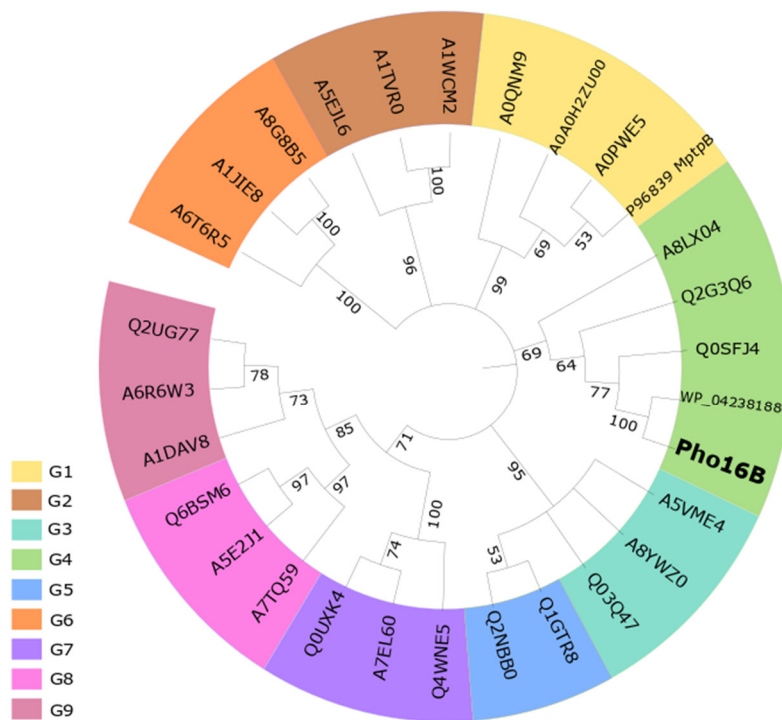


Figure 3. Neighbor-joining phylogenetic tree showing the position of Pho16B. The numbers at the nodes indicate levels of bootstrap support (range from 0 to 100) and were based on 500 replicates. Names

correspond to the groups (G) described by Beresford et al. (2010) and their corresponding UniProtKB codes. Pho16B, this study; WP_042381880 (closest related PTP phosphatase from *Streptacidiphilus melanogenes*); P96839 (MptpB *Mycobacterium tuberculosis*); G1: A0QNM9 (*Mycobacterium smegmatis*), A0PWE5 (*Mycobacterium ulcerans*), A0A0H2ZU00 (*Mycobacterium avium*); G2: A1TVR0 (*Acidovorax avenae*), A1WCM2 (*Acidovorax* sp.), A5EJL6 (*Bradyrhizobium* sp.); G3: A5VME4 (*Lactobacillus reuteri*), A8YWZ0 (*Lactobacillus helveticus*), Q03Q47 (*Lactobacillus brevis*); G4: A8LX04 (*Salinispora arenicola*), Q0SFJ4 (*Rhodococcus jostii*), Q2G3Q6 (*Novosphingobium aromaticivorans*); G5: Q1GTR8 (*Sphingopyxis alaskensis*), Q2NBB0 (*Erythrobacter litoralis*); G6: A1JIE8 (*Yersinia enterocolitica*), A6T6R5 (*Klebsiella pneumoniae*), A8G8B5 (*Serratia proteamaculans*); G7: A7EL60 (*Sclerotinia sclerotiorum*), Q0UXK4 (*Phaeosphaeria nodorum*), Q4WNE5 (*Neosartorya fumigata*); G8: A5E2J1 (*Lodderomyces elongisporus*), A7TQ59 (*Vanderwaltozyma polyspora*), Q6BSM6 (*Debaryomyces hansenii*); G9: A6R6W3 (*Ajellomyces capsulatus*), Q2UG77 (*Aspergillus oryzae*), A1DAV8 (*Neosartorya fischeri*).

2.2. Tertiary Structure Prediction of Pho16B

In addition to the sequence comparisons, we further predicted the theoretical 3D model of Pho16B, to determine its closest structural relative. This analysis was performed by using the I-TASSER tools [31]. For the modeling, I-TASSER initially uses different threading programs generating tens of thousands of template alignments. Only the templates of the highest significance in the threading alignments were used. The significance of the alignments was measured by a normalized Z-score. Z-scores > 1 mean a good alignment and vice versa (Table S2).

Employment of I-TASSER suite resulted in the prediction of five models for Pho16B. Two of the predicted models (M1 and M2) showed C-score values higher than -1.5 (-1.22 and -1.36, respectively). Therefore, the model M1 was selected as the most plausible model for Pho16B (Figure S1). The model M1 was related to the Protein Data Bank (PDB) entry 1YWF, that corresponds to the crystal structure of *M. tuberculosis* PTP (PtpB). It is assumed that a C-score higher than -1.5 confirms a correct global topology [31]. As the native structure of Pho16B is not known, the quality of the modeling prediction is determined by calculating the distance between the predicted model and published native structures. In our case, I-TASSER predicted the quality of the model by calculating the TM-score. A TM-score with a value of 0.56 was calculated, indicating a very similar fold of the Pho16B and the reference protein PtpB [32]. The analysis of the occurrence and distribution of clinically relevant bacterial virulence genes across natural (non-human) environments has demonstrated the presence of transcribing homologs of several virulence genes in those habitats. Consequently, a common ancestral origin between the environmental genes and their virulence-associated counterparts has been established [3]. Pho16B is the first reported environmental PTP with phytase activity, carrying a homologous molecular signature of virulence gene-related products (ALPs). Therefore, this study provides insights into the structural similarities of environmental PTPs to other known phosphatases related to virulence factors of human pathogens.

2.3. Optimum Temperature and pH of Pho16B

The characterization of the biochemical properties of Pho16B revealed its optimal activity to occur at a temperature of 45 °C. More than 50% of its activity was lost at temperatures ≥ 50 °C (Figure 4A). In general, phytases show activity within a wide range of temperatures, and some previously reported bacterial representatives are affected by temperature in a similar way as Pho16B. The phytase of *B. bacteriovorus* Bd1204 exhibits optimal activity at a slightly higher temperature than Pho16B, but also loses most of its activity (40%) at temperatures higher than 50 °C [30].

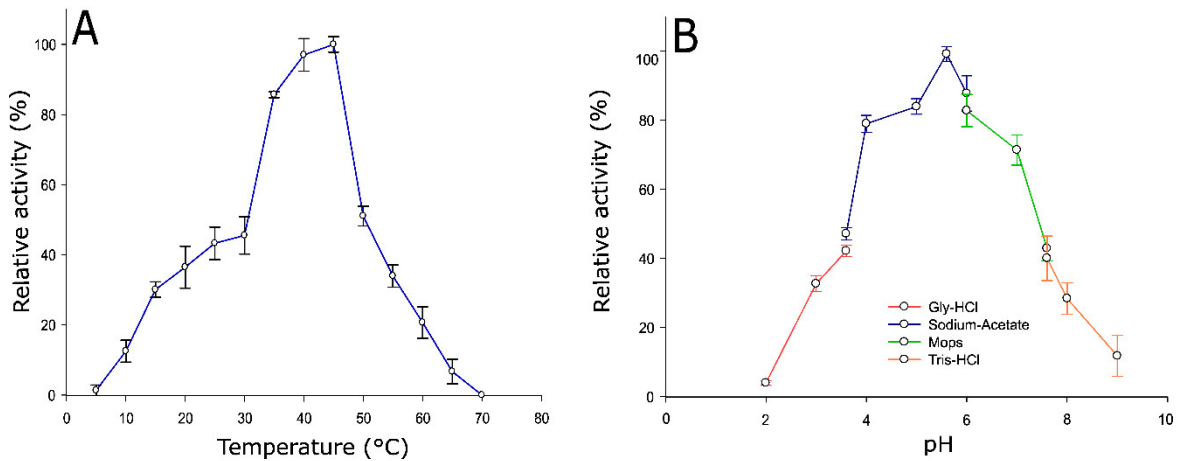


Figure 4. Effect of temperature and pH on Pho16B activity. **(A)** Temperature profile of Pho16B enzymatic activity. **(B)** pH profile of Pho16B enzymatic activity. All measurements were performed in triplicate. Specific activity values are expressed as percentages of the highest relative activity: 8.78 and 4.3 U/mg, for A and B, respectively.

The optimal pH of Pho16B enzymatic activity (5.6) was detected by measurements across a set of overlapping buffer systems ranging from pH 2 to 9. At pH 4.0 and neutral pH, Pho16B retained 80% and 70% of the activity, respectively. At slightly alkaline pH value of 7.5, activity was reduced to 50% (Figure 4B). The activities of other reported PTP phytases show a strong reduction at this pH level, i.e., the enzyme phyAme from *Megasphaera elsdenii* loses its activity almost completely at pH 7.0 [33]. The optimal activity and higher capacity of enzymatic activity retention of Pho16B at low pH match with the predicted extracellular localization of the enzyme at its natural acidic forest soil habitat, which has a soil pH of 3.5 [34,35]. This is also consistent with the cysteine-based catalytic mechanism of MptpB and other PTPs, whose optimal catalytic activities are at acidic pH values [36].

2.4. Pho16B Converts a Broad Range of Substrates

The substrate specificity of Pho16B was tested using ten different substrates (Figure 5). Pho16B showed activity in the presence of all tested compounds. The highest relative activity was detected with naphthyl phosphate, and the lowest with NADP as substrate. The relative activity with InsP₆ was approximately 30% of that with naphthyl phosphate (Figure 5). This indicated that InsP₆ is not the primary substrate of Pho16B under the tested conditions. Nevertheless, to our knowledge, this is the first reported example of an environmental PTP with activity towards InsP₆. We previously demonstrated the phytase activity of environmental phosphatases such as alkaline phosphatases and acid phosphatases [28]. Many phosphatases are recognized as promiscuous types of enzymes [37]. The broad substrate spectrum indicated that Pho16B belongs also to this type.

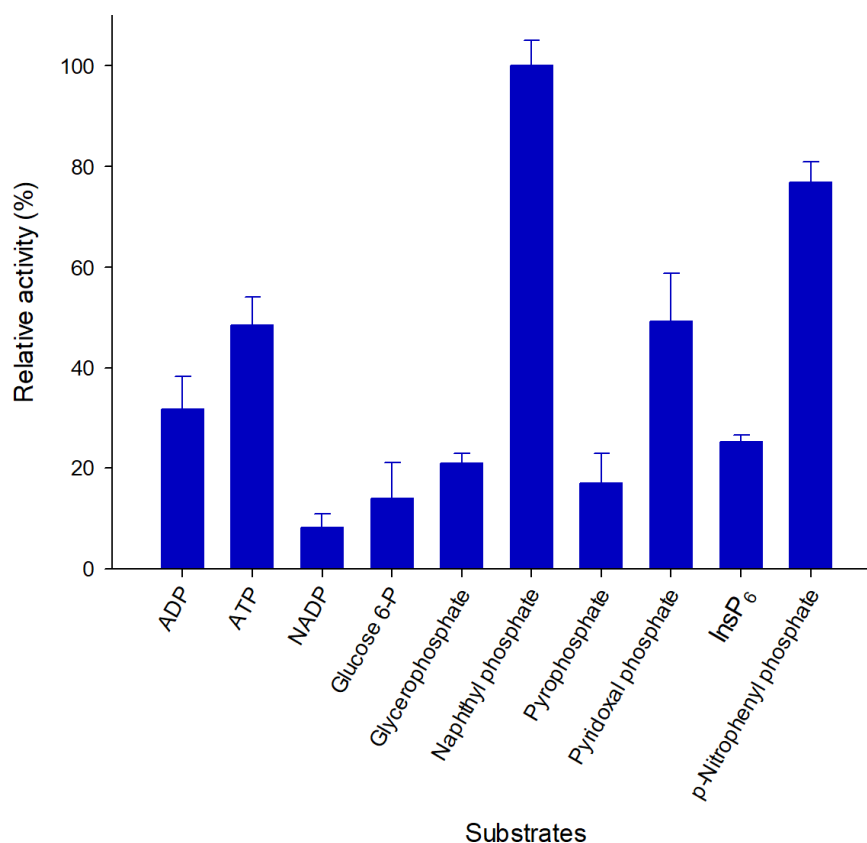


Figure 5. Substrate specificity of Pho16B. Relative activity of Pho16B was measured at 10 mM substrate concentration. All measurements were performed in triplicate and under optimal pH and temperature conditions for enzyme activity. Specific activity values are expressed as percentages of the highest relative activity (13.89 U/mg).

The kinetic parameters of Pho16B were determined using the purified protein and InsP₆, naphthyl phosphate, and *p*-nitrophenylphosphate as substrates under optimal pH and temperature conditions (pH 5.6 and 45 °C; Table 1). Pho16B shows a higher affinity and catalytic efficiency for naphthyl phosphate than for InsP₆. The protein tyrosine phosphatase (PhyAsr) from *Selenomonas ruminantium*, which is one of the few characterized and reported (PTP)-like phytases, possesses similar affinity (K_m) and catalytic efficiency (K_{cat}/K_m) towards InsP₆ under high ionic strength conditions (1 mM and 163 mM⁻¹s⁻¹, respectively) [38]. The protein tyrosine phosphatase of *Yersinia Yop51*, which is involved in pathogenesis, displays affinity values almost three times lower than Pho16B with *p*-nitrophenylphosphate as substrate (K_m 2.90 mM). However, the turnover number of Yop51 (1235 ± 36 s⁻¹) is significantly higher than the one of Pho16B [39]. Protein phosphatases, including phytases, possess a wide range of catalytic traits. In this sense, the analyzed kinetic characteristics of Pho16B are in the normal range compared with the currently reported phytases [14,18,39]. However, since Pho16B is the first reported environmental PTP, it is not possible to compare its kinetic parameters with those from phytases of the same type.

Table 1. Kinetic values of Pho16B under optimal pH and temperature conditions. All measurements were performed in triplicate.

Substrate	K_m (mM)	k_{cat} (min ⁻¹)	K_{cat}/K_m (s ⁻¹ mM ⁻¹)
InsP ₆	1.290 ± 0.38	5.48 ± 0.7	70.43 ± 4.4
Naphthyl phosphate	0.966 ± 0.18	14.64 ± 1.52	238.73 ± 14.52
<i>p</i> -Nitrophenylphosphate	1.026 ± 0.14	19.73 ± 2.8	316.89 ± 32.1

2.5. Effect of Additives on Pho16B Activity

The effects of various additives on Pho16B enzyme activity are summarized in Figure 6. Co^{2+} , Cu^{2+} , and Fe^{2+} inhibit enzyme activity by more than 50%. In the presence of Al^{3+} , no activity was detected. It has been previously demonstrated that the phosphatase activity of soil-derived enzymes might be inhibited in the presence of aluminum hydroxides [40]. Moreover, Cu^{2+} and Fe^{2+} have been reported as strong inhibitors of several phosphatases and phytases from different organisms, such as *S. ruminantium* or *Klebsiella terrigena* [41,42]. These proteins were also strongly inhibited in the presence of Zn^{2+} , which only has a minor effect on the activity of Pho16B. The presence of Co^{2+} reduced the relative activity of Pho16B by 80%. By contrast, other phosphatases, i.e., the enzymes derived from bovine rumen bacterium *Mitsuokella jalaludinii*, are not affected by the presence of Co^{2+} [42]. The glucose-1-phosphatase with phytase activity (AgpP) from *Pantoea agglomerans* possesses an unusual metal ion activation. Ca^{2+} , Mg^{2+} , and Mn^{2+} enhance the activity of this enzyme up to 200% [43]. The activity of Pho16B was not affected by the addition of Ca^{2+} or Mg^{2+} , but the presence of Mn^{2+} slightly increased the enzymatic activity of Pho16B.

The addition of dithiothreitol (DTT) had a very strong inhibitory effect, depleting 80% of Pho16B activity. Other previously reported phosphatases did not show significant losses of activity in the presence of this reagent [44,45]. It has been shown that DTT can reduce protein function by disrupting disulfide bonds of the proteins or acting as a metal ion chelator [46]. Another possibility is that DTT could inhibit Pho16B activity by competing with its substrates. It has been proposed that DTT can interact with the catalytic domain of the enzymes by steric hindrance via hydrogen interactions with amino acid residues [47]. Wolframate and oxalate also reduce the activity of Pho16B to less than 60%. Oxalate is known as an inhibitor of acid phosphatase activity [48], and wolframate as inhibitor of phosphatase activity of PTPs from plants [49].

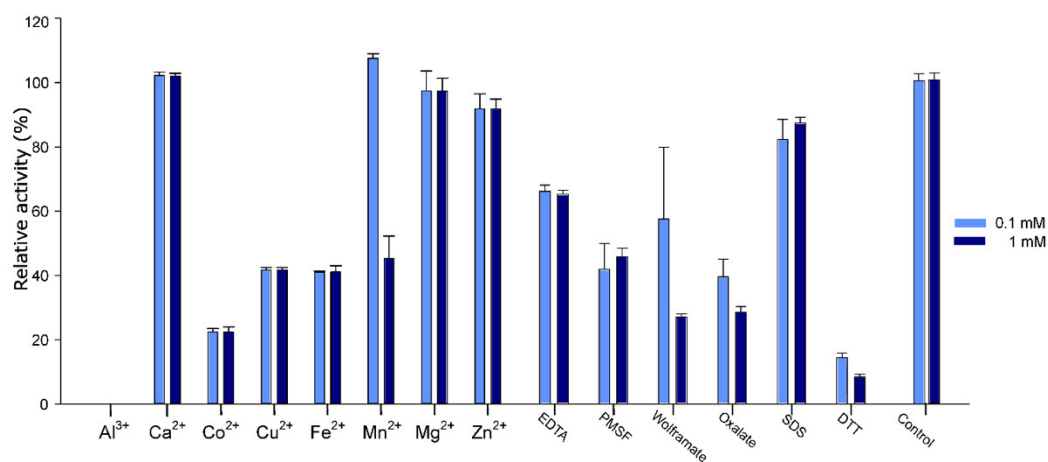


Figure 6. Effect of different concentrations of metal ions and inhibitors on the activity of Pho16B. All measurements were performed in triplicate and under optimal pH and temperature conditions for the enzyme. Specific activity values expressed as percentages of the control reactions (no additions) 8.2 U/mg for both concentrations.

3. Materials and Methods

3.1. Sampling, Metagenomic Library Construction, and Function-Based Screening

The enzyme Pho16B originated from an Arenosol soil A horizon sample (SEW2) with a pH of 3.46, which was collected from a beech forest site within the Schorfheide-Chorin biosphere reserve in Germany. Collection of the sample was performed as previously described by Kaiser et al. (2016) [50]. The metagenomic library (SEW2) comprised 135,240 *E. coli* clones and was constructed using the plasmid pCR-XL-TOPO as vector (Invitrogen GmbH, Karlsruhe, Germany) and the method described by Nacke et al. [35]. The previous screening of the metagenomic library was performed by following the method of Villamizar et al.. Phosphatase/phytase positive library-bearing *E. coli* clones growing on modified Sperber

minimal medium, using InsP_6 as phosphorous source and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) as indicator, turn from white to dark blue within 48 h [51].

3.2. Molecular Analysis

The insert sequence of plasmid pLP16 (7972 bp), which was isolated from a metagenomic library clone showing phosphatase activity [28], was subjected to sequence analysis. Open reading frame (ORF) prediction was performed using the ORF finder tool provided by the NCBI and the ARTEMIS program [52,53]. Since the metagenomic library was constructed from metagenomic DNA fragments and cloned into the pCR-XL-TOPO vector, all genes predicted for both strands were analyzed, in order to identify the putative gene or genes responsible for the activity on the indicator plate. The results were verified and improved manually by using criteria such as the presence of a ribosome-binding site, GC frame plot analysis, and similarity to known genes. Amino acid sequences deduced from the ORFs were examined for similarities with known protein families and domains by performing searches against the InterPro collection of protein signature database and the Conserved Domain Database (CDD). Next, a search of phosphatase signatures was performed by using a local version of PhosphaBase [54–56]. Signal peptide prediction was performed using SIGNALP 4.0 [57]. Additionally, the taxonomic classification of the complete DNA insert was performed by using the software KAIJU [58]. Similarity searches of candidate gene product of *pho16B*, were performed by using the NCBI databases non-redundant sequences (nr) and metagenomic proteins (env_nr), and a Basic Local Alignment Search for proteins (BLASTP). A second search against metagenomic data was performed by using the metagenomics platform of the European Bioinformatics Institute [59,60].

Multiple sequence alignments of Pho16B encoded by gene *pho16B* and related phosphatases, representing the phosphatase groups previously defined by Beresford et al. [6], were performed using MUSCLE [61]. A phylogenetic consensus tree was calculated using the neighbor-joining method with MEGA X and 500 bootstrap replicates [62]. The tree was visualized using iTOL v3 [63]. The evolutionary distances were calculated using the number of differences method [3].

A prediction of the tertiary structure of protein Pho16B was performed by employing the I-TASSER software suite [64,65]. The quality of models generated using I-TASSER is based on two major criteria: the confidence score (C-score) and the template modeling score (TM-score). I-TASSER generated five models. The models were ranked based on the C-score (confidence score). The C-scores are calculated on the basis of the statistical significance of the threading profile–profile alignment, as well as structure convergence of the assembly simulations. The C-score ranged from -5 to 2. A high C-score value indicates a model with greater confidence [32].

The TM-score (the template modeling score) addresses the structural similarity of two protein models. This score can solve some difficulties of the commonly used metrics, such as the root-mean-square deviation (RMSD). The TM-score measures the global fold similarity. Moreover, TM-score is less sensitive to local structural variations. Another advantage of this measurement is that the magnitude of the TM-score for random structure pairs is length-independent. The TM-score has a value range of 0 to 1, whereby the value 1 indicates a perfect match between two structures [32]. By calculating the TM-score, we obtained an estimation about the structural similarity between the predicted model of Pho16B and published native or experimentally determined structures. Values close to 0.5 indicate a model of correct topology. In this study, the model with the highest confidence score (C-score) was selected as the best predicted optimized 3D modeling structure.

3.3. Biochemical Characterization of Pho16B

The ORF *pho16B* was cloned into plasmid pBAD202/D-TOPO (Thermo Fisher Scientific GmbH, Schwerte, Germany). In this way, sequences encoding the His₆ and thioredoxin tags were added to the N terminus of Pho16B. Plasmid DNA containing the insert cloned in the correct orientation was used for transforming *E. coli* LMG194 cells. Transformants were grown on Luria-Bertani (LB) agar plates [66] supplemented with kanamycin (50 $\mu\text{g}/\mu\text{L}$) and incubated at 37 °C. A colony of *E. coli* LMG194 harboring the *pho16B*-pBAD202 construct was used to inoculate 500 mL M9 minimal salts medium [67] supplemented with 50 $\mu\text{g}/\mu\text{L}$ kanamycin and 2% glycerol. The culture was incubated

under shaking at 90 rpm by using a New Brunswick Innova 44 incubator-shaker (Eppendorf AG, Hamburg, Germany) at 30 °C. Protein expression was induced at an OD₆₀₀ of 0.8 using L-arabinose (end concentration 0.02%) and incubated for 16 h.

The cells were harvested by centrifugation for 30 min at 4 °C and 8000 rpm (Sorvall® RC6 centrifuge, rotor SLA 3000, Thermo Fisher Scientific). The resulting cell pellets were suspended in 10 mL 50 mM Tris, 250 mM NaCl buffer. Mechanical cell disruption was performed using a French press (1.38×10^8 Pa; Thermo Fisher Scientific). Subsequently, the extract was cleared by centrifugation for 0.5 h at 4 °C and 15,000 rpm (Sorvall® RC6 centrifuge with rotor SS 35, Thermo Fisher Scientific). The crude extract was filtered twice using 0.45 µm and 0.2 µm filters (Sarstedt, Nümbrecht, Germany). In order to purify the His₆-tagged protein, the Protino® Ni-TED 2000 purification kit was used as recommended by the manufacturer (Macherey and Nagel, Düren, Germany) with modifications. The equilibration of the columns and the washing steps were performed with 50 mM Tris buffer (pH 8.0) containing 250 mM NaCl, followed by three elution steps with 50 mM Tris, 250 mM NaCl, and 250 mM imidazole.

To concentrate the protein, eliminate free P traces, remove imidazole, and reduce NaCl concentration, four subsequent rounds of ultrafiltration using 50 mM Tris buffer (pH 8.0) were conducted by using Vivaspin concentrators, as recommended by the manufacturer (exclusion limit 10 kDa; Sartorius AG, Göttingen, Germany). To remove the thioredoxin tag, fusion proteins were digested by a light chain enterokinase (NEB, Ipswich, MA, USA) (1.5 U/mg fusion protein) at 37 °C for 16 h in buffer (20 mM Tris-HCl, 2 mM CaCl₂, 50 mM NaCl, pH 8.0). Further purification was performed via hydrophobic interaction chromatography by using a 15PHE 4.6/100PE Tricorn high performance column (bed volume 1.7 mL, flowrate 2 mL/min) and an ÄKTA FPLC system (GE Healthcare, Uppsala, Sweden). Elution was achieved by a linear decreasing NaCl gradient from 1 M to 0 M in 50 mM Tris-HCl (pH 8).

Phosphatase activity was determined at 355 nm by detecting the release of inorganic phosphorous according to the ammonium molybdate method developed by Heinonen and Lahti with modifications [68,69]. The purified enzyme solution (10 µL) was pre-incubated for 3 min at 40 °C in 380 µL of 50 mM sodium acetate buffer (pH 5). Subsequently, 10 µL of 100 mM phytic acid dipotassium salt (Sigma-Aldrich, Munich, Germany) was added, and the mixture incubated for 30 min at 40 °C. To stop the reaction, 1.5 mL of freshly prepared AAM solution (acetone–5 N H₂SO₄–10 mM ammonium molybdate) and 100 µL of 1 M citric acid were added. Samples were measured against blanks prepared by adding AAM solution prior to the addition of enzyme. The absorbance (355 nm) was measured using the Ultrospec® 3300 pro (Amersham plc, Little Chalfont, United Kingdom). All measurements were performed in triplicate. To calculate the enzyme activity, a calibration curve was generated in the range of 5 to 600 nmol phosphate. One activity unit (U) represented the release of 1 nmol phosphate release per min.

The influence of temperature on enzymatic activity was determined via the above-described standard phytase assay. The enzymatic activity was evaluated in a temperature range between 5 and 70 °C by using a temperature-adjusted buffer (50 mM sodium acetate, pH 6). In order to analyze the pH dependence of enzyme activity, the following overlapping buffers were prepared as described by Gomori 1955 [70]: 50 mM glycine-HCl (pH 2.0, 3.0, and 3.6), sodium acetate (pH 3.6, 4.0, 5.0, 5.6, and 6.0), MOPS (pH 6.0, 7.0, and 7.6), Tris-HCl (pH 7.6, 8.0, and 9.0), and glycine-NaOH (pH 9.0).

The substrate specificity was determined using the standard phytase assay under optimal temperature and pH. Ten different substrates (ADP, ATP, NADP, glucose-6-phosphate, glycerophosphate, pyridoxal phosphate, pyrophosphate, naphthyl phosphate, *p*-nitrophenylphosphate, and InsP₆) were tested at 10 mM. Furthermore, the effect of cations (Al³⁺, Ca²⁺, Cu²⁺, Co²⁺, Fe²⁺, Mg²⁺, Mn²⁺, and Zn²⁺) and the potential inhibitors ethylenediaminetetraacetic acid (EDTA), phenylmethylsulfonyl fluoride (PSMF), wolframate, oxalate, sodium dodecyl sulfate (SDS), and dithiothreitol (DTT), at concentrations of 0.1 and 1 mM, were analyzed.

Kinetic parameters K_m and V_{max} for Pho16B were calculated from the Michaelis–Menten equation fitted to a non-linear, least-squares regression by using the kinetics module of the program SigmaPlot 12.0 (Systat Software, Inc., San Jose, CA, USA). All measurements were performed under optimal pH

and temperature conditions using InsP_6 , *p*-nitrophenylphosphate and naphthyl phosphate as substrates.

3.4. Accession Number

The nucleotide sequence of the pLP16 insert sequence has been submitted to the National Center for Biotechnology Information (NCBI) GenBank under accession number KY931681.

4. Conclusions

Although PTPs have been relatively well studied, the diversity, role, and characteristics of environmental PTPs remain unknown. All previously described and characterized PTPs were derived from individual microorganisms. Here, for the first time, we characterized an environmentally and metagenome-derived PTP (Pho16B). Pho16B exhibits the characteristic motif of ALPs, which are associated with microbial pathogenesis. At the same time, Pho16B is the first environmentally derived PTP capable of using InsP_6 as substrate and, thus, the first PTP-phy which does not originate from an isolated microorganism. PTP-phys were described one decade ago, but very little is known about their biological role [21]. The capability of Pho16B to use InsP_6 as substrate under acidic conditions exposes the relevance of soil as a source of interesting new phytases. Environmental phytases have the potential to be used to solve problems such as eutrophication, associated with the presence of phytic acid in subterranean waters and other bodies of water. New phytases derived from natural environments can be used in agriculture for the rational design of pest inhibitors for crop protection. Similar approaches have been successfully used by analyzing PTPs associated with human diseases, such as diabetes and cancer [71]. Moreover, the high degree of sequence conservation between functional environmental PTPs like Pho16B, and enzymes associated to pathogenicity processes, provides support to the idea of environmental homologous genes as precursors of virulence genes found in clinically relevant bacteria [3]. Therefore, the information obtained from new environmental PTPs provides new valuable insights into the origin of this type of molecule.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1. Protein sequence similarities of deduced gene products encoded by the insert of pLP16. Table S2. Rank of templates representing the top ten threading PDB templates used by I-TASSER. Figure S1 (A) 3D structure of M1 for the Pho16B protein, predicted by I-TASSER, calculated C score -1.22 . (B) Amino acids involved in the P-loop in Pho16B.

Author Contributions: R.D., H.N. developed the concept. G.A.C.V., L.G. and K.F. performed microbiological and molecular experiments. G.A.C.V., L.T. performed in silico analysis of the sequence. G.A.C.V. drafted the manuscript. R.D. supervised the work. H.N., R.D. and L.T. were involved in revision of the manuscript and figures. All authors interpreted the results, contributed to the writing of the final version of the manuscript and approved submission.

Funding: We thank the Deutsche Akademische Austauschdienst (DAAD) and Colciencias Colombia for the financial support to Genis Andrés Castillo Villamizar in the frame of the fellowship program ALECOL. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments: We thank Silja Brady and Mechthild Bömeke for providing technical assistance. We acknowledge support by the Open Access Publication Funds of the University of Göttingen.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Pallen, M.J.; Wren, B.W. Bacterial pathogenomics. *Nature* **2007**, *449*, 835–842, doi:10.1038/nature06248.
2. Daniel, R. The metagenomics of soil. *Nat. Rev. Microbiol.* **2005**, *3*, 470–478, doi:10.1038/nrmicro1160.

3. Søborg, D.A.; Hendriksen, N.B.; Kilian, M.; Christensen, J.H.; Kroer, N. Bacterial human virulence genes across diverse habitats as assessed by in silico analysis of environmental metagenomes. *Front. Microbiol.* **2016**, *7*, 1712, doi:10.3389/fmicb.2016.01712.
4. Tiganis, T.; Bennett, A.M. Protein tyrosine phosphatase function: The substrate perspective. *Biochem. J.* **2007**, *402*, 1–15, doi:10.1042/bj20061548
5. Tonks, N.K. Protein tyrosine phosphatases: From genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 833–846, doi:10.1038/nrm2039.
6. Beresford, N.J.; Saville, C.; Bennett, H.J.; Roberts, I.S.; Tabernero, L. A new family of phosphoinositide phosphatases in microorganisms: Identification and biochemical analysis. *BMC Genom.* **2010**, *11*, 457, doi:10.1186/1471-2164-11-457.
7. Beresford, N.; Patel, S.; Armstrong, J.; Szöör, B.; Fordham-Skelton, A.P.; Tabernero, L. Mptpb, a virulence factor from *Mycobacterium tuberculosis*, exhibits triple-specificity phosphatase activity. *Biochem. J.* **2007**, *406*, 13–18, doi:10.1042/BJ20070670.
8. Puhl, A.A.; Greiner, R.; Selinger, L.B. Kinetics, substrate specificity, and stereospecificity of two new protein tyrosine phosphatase-like inositol polyphosphatases from *Selenomonas lacticifex*. *Biochem. Cell Biol.* **2008**, *86*, 322–330, doi:10.1139/O08-095.
9. Weber, S.; Stirnimann, C.U.; Wieser, M.; Frey, D.; Meier, R.; Engelhardt, S.; Li, X.; Capitani, G.; Kammerer, R.A.; Hilbi, H. A type IV translocated *Legionella* cysteine phytase counteracts intracellular growth restriction by phytate. *J. Biol. Chem.* **2014**, *289*, 34175–34188, doi:10.1074/jbc.M114.592568.
10. Irvine, R.F.; Schell, M.J. Back in the water: The return of the inositol phosphates. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 327–338, doi:10.1038/35073015.
11. Carr, J.P.; Lewsey, M.G.; Palukaitis, P. Signaling in induced resistance. In *Advances in Virus Research*; Carr, J.P., Loebenstein, G., Eds.; Academic Press: Cambridge, MA, USA, 2010; Volume 76, pp. 57–121.
12. Graf, E.; Empson, K.L.; Eaton, J.W. Phytic acid. A natural antioxidant. *J. Biol. Chem.* **1987**, *262*, 11647–11650.
13. Singh, R.P.; Agarwal, R. Prostate cancer and inositol hexaphosphate: Efficacy and mechanisms. *Anticancer Res.* **2005**, *25*, 2891–2903
14. Bruder, L.M.; Gruninger, R.J.; Cleland, C.P.; Mosimann, S.C. Bacterial PhyA protein-tyrosine phosphatase-like myo-inositol phosphatases in complex with the Ins(1,3,4,5)P₄ and Ins(1,4,5)P₃ second messengers. *J. Biol. Chem.* **2017**, *292*, 17302–17311, doi:10.1074/jbc.M117.787853.
15. Sasakawa, N.; Sharif, M.; Hanley, M.R. Metabolism and biological activities of inositol pentakisphosphate and inositol hexakisphosphate. *Biochem. Pharmacol.* **1995**, *50*, 137–146, doi:10.1016/0006-2952(95)00059-9.
16. Mullaney, E.J.; Daly, C.B.; Ullah, A.H.J. Advances in phytase research. In *Advances in Applied Microbiology*; Academic Press: Cambridge, MA, USA, 2000; Volume 47, pp. 157–199, ISBN 978-0-12-002647-0.
17. Gerke, J. Phytate (inositol hexakisphosphate) in soil and phosphate acquisition from inositol phosphates by higher plants. A review. *Plants* **2015**, *4*, 253–266, doi:10.3390/plants4020253.
18. Mullaney, E.J.; Ullah, A.H.J. The term phytase comprises several different classes of enzymes. *Biochem. Biophys. Res. Commun.* **2003**, *312*, 179–184, doi:10.1016/j.bbrc.2003.09.176.
19. Haefner, S.; Knietzsch, A.; Scholten, E.; Braun, J.; Lohscheidt, M.; Zelder, O. Biotechnological production and applications of phytases. *Appl. Microbiol. Biotechnol.* **2005**, *68*, 588–597, doi:10.1007/s00253-005-0005-y.
20. Ushasree, M.V.; Vidya, J.; Pandey, A. Other enzymes: Phytases. In *Current Developments in Biotechnology and Bioengineering*; Pandey, A., Negi, S., Soccol, C.R., Eds.; Elsevier: Amsterdam, The Netherlands, 2017; pp. 309–333.
21. Puhl, A.A.; Gruninger, R.J.; Greiner, R.; Janzen, T.W.; Mosimann, S.C.; Selinger, L.B. Kinetic and structural analysis of a bacterial protein tyrosine phosphatase-like myo-inositol polyphosphatase. *Protein Sci.* **2007**, *16*, 1368–1378, doi:10.1110/ps.062738307.
22. Lei, X.G.; Weaver, J.D.; Mullaney, E.; Ullah, A.H.; Azain, M.J. Phytase, a new life for an “old” enzyme. *Annu. Rev. Anim. Biosci.* **2013**, *1*, 283–309, doi:10.1146/annurev-animal-031412-103717.
23. Gruninger, R.J.; Dobing, S.; Smith, A.D.; Bruder, L.M.; Selinger, L.B.; Wieden, H.J.; Mosimann, S.C. Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases. *J. Biol. Chem.* **2012**, *287*, 9722–9730, doi:10.1074/jbc.M111.309872.

24. Farias, N.; Almeida, I.; Meneses, C. New bacterial phytase through metagenomic prospecting. *Molecules* **2018**, *23*, 448, doi:10.3390/Molecules23020448.
25. Berini, F.; Casciello, C.; Marcone, G.L.; Marinelli, F. Metagenomics: Novel enzymes from non-culturable microbes. *FEMS Microbiol. Lett.* **2017**, *364*, fnx211, doi:10.1093/femsle/fnx211.
26. Tan, H.; Wu, X.; Xie, L.; Huang, Z.; Peng, W.; Gan, B. Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 2225–2241, doi:10.1007/s00253-015-7097-9.
27. Tan, H.; Mooij, M.J.; Barret, M.; Hegarty, P.M.; Harington, C.; Dobson, A.D.; O’Gara, F. Identification of novel phytase genes from an agricultural soil-derived metagenome. *J. Microbiol. Biotechnol.* **2014**, *24*, 113–118, doi:10.4014/jmb.1307.07007.
28. Castillo Villamizar, G.A.; Nacke, H.; Böhning, M.; Herz, K.; Daniel, R. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio* **2019**, *10*, e01966-18. doi:10.1128/mBio.01966-18.
29. Cho, S.-H.; Han, J.-H.; Ko, H.-Y.; Kim, S.B. *Streptacidiphilus anmyonensis* sp. Nov., *Streptacidiphilus rugosus* sp. Nov. and *Streptacidiphilus melanogenes* sp. Nov., acidophilic actinobacteria isolated from pinus soils. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 1566–1570, doi:10.1099/ijs.0.65480-0.
30. Gruninger, R.J.; Thibault, J.; Capeness, M.J.; Till, R.; Mosimann, S.C.; Sockett, R.E.; Selinger, B.L.; Lovering, A.L. Structural and biochemical analysis of a unique phosphatase from *Bdellovibrio bacteriovorus* reveals its structural and functional relationship with the protein tyrosine phosphatase class of phytase. *PLoS ONE* **2014**, *9*, e94403, doi:10.1371/journal.pone.0094403.
31. Yang, J.; Zhang, Y. Protein structure and function prediction using I-TASSER. *Curr. Protoc. Bioinform.* **2015**, *52*, 5.8.1–5.8.15, doi:10.1002/0471250953.bi0508s52.
32. Roy, A.; Kucukural, A.; Zhang, Y. I-tasser: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738, doi:10.1038/nprot.2010.5.
33. Puhl, A.A.; Greiner, R.; Selinger, L.B. Stereospecificity of myo-inositol hexakisphosphate hydrolysis by a protein tyrosine phosphatase-like inositol polyphosphatase from *Megasphaera elsdenii*. *Appl. Microbiol. Biotechnol.* **2009**, *82*, 95–103, doi:10.1007/s00253-008-1734-5.
34. Anné, J.; Vrancken, K.; Van Mellaert, L.; Van Impe, J.; Bernaerts, K. Protein secretion biotechnology in gram-positive bacteria with special emphasis on *Streptomyces lividans*. *Biochim. Biophys. Acta* **2014**, *1843*, 1750-1761. doi: 10.1016/j.bbamcr.2013.12.023.
35. Nacke, H.; Will, C.; Herzog, S.; Nowka, B.; Engelhaupt, M.; Daniel, R. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol. Ecol.* **2011**, *78*, 188–201. doi:10.1111/j.1574-6941.2011.01088.x.
36. Heneberg, P.; Draber, P. Regulation of cys-based protein tyrosine phosphatases via reactive oxygen and nitrogen species in mast cells and basophils. *Curr. Med. Chem.* **2005**, *12*, 1859–1871, doi:10.2174/0929867054546636.
37. Huang, H.; Pandya, C.; Liu, C.; Al-Obaidi, N.F.; Wang, M.; Zheng, L.; Toews Keating, S.; Aono, M.; Love, J.D.; Evans, B.; et al. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E1974–E1983, doi:10.1073/pnas.1423570112.
38. Gruninger, R.J.; Selinger, L.B.; Mosimann, S.C. Effect of ionic strength and oxidation on the p-loop conformation of the protein tyrosine phosphatase-like phytase, phyasr. *FEBS J.* **2008**, *275*, 3783–3792, doi:10.1111/j.1742-4658.2008.06524.x.
39. Zhang, Z.Y.; Clemens, J.C.; Schubert, H.L.; Stuckey, J.A.; Fischer, M.W.; Hume, D.M.; Saper, M.A.; Dixon, J.E. Expression, purification, and physicochemical characterization of a recombinant *Yersinia* protein tyrosine phosphatase. *J. Biol. Chem.* **1992**, *267*, 23759–23766.
40. George, T.S.; Quiquampoix, H.; Simpson, R.J.; Richardson, A.E. Interactions between phytases and soil constituents: Implications for the hydrolysis of inositol phosphates. In *Inositol Phosphates, Linking Agriculture and the Environment*; Turner, B.L., Richardson, A.E., Mullaney, E.J., Eds.; CABI: Wallingford, UK, 2007; Volume 1, pp. 221–241.
41. Bekalu, Z.E.; Madsen, C.K.; Dionisio, G.; Brinch-Pedersen, H. *Aspergillus ficuum* phytase activity is inhibited by cereal grain components. *PLoS ONE* **2017**, *12*, e0176838, doi:10.1371/journal.pone.0176838.

42. Lan, G.Q.; Abdullah, N.; Jalaludin, S.; Ho, Y.W. Purification and characterization of a phytase from *Mitsuokella jalaludinii*, a bovine rumen bacterium. *Afr. J. Biotechnol.* **2011**, *10*, 12766–12776, doi:10.5897/AJB11.294.
43. Suleimanova, A.D.; Beinhauer, A.; Valeeva, L.R.; Chastukhina, I.B.; Balaban, N.P.; Shakirov, E.V.; Greiner, R.; Sharipova, M.R. Novel glucose-1-phosphatase with high phytase activity and unusual metal ion activation from soil bacterium *Pantoea* sp. Strain 3.5.1. *Appl. Environ. Microbiol.* **2015**, *81*, 6790–6799, doi:10.1128/AEM.01384-15.
44. Monteiro, P.S.; Guimarães, V.M.; de Melo, R.R.; de Rezende, S.T. Isolation of a thermostable acid phytase from *Aspergillus niger* ufv-1 with strong proteolysis resistance. *Braz. J. Microbiol.* **2015**, *46*, 251–260, doi:10.1590/s1517-838220120037.
45. Xing, Raza, A.; Löfgren, S.; Fernando, M.R.; Ho, Y.-S.; Lou, M.F. Low molecular weight protein tyrosine phosphatase (LMW-PTP) and its possible physiological functions of redox signaling in the eye lens. *Biochim. Biophys. Acta* **2007**, *1774*, 545–555, doi:10.1016/j.bbapap.2007.03.001.
46. Alves, M.P.; Salgado, R.L.; Eller, M.R.; Vidigal, P.M.P.; Fernandes de Carvalho, A. Characterization of a heat-resistant extracellular protease from *Pseudomonas fluorescens* 07a shows that low temperature treatments are more effective in deactivating its proteolytic activity. *J. Dairy Sci.* **2016**, *99*, 7842–7851, doi:10.3168/jds.2016-11236.
47. Alliegro, M.C. Effects of dithiothreitol on protein activity unrelated to thiol–disulfide exchange: For consideration in the analysis of protein function with cleland’s reagent. *Anal. Biochem.* **2000**, *282*, 102–106, doi:10.1006/abio.2000.4557.
48. Igarashi, M.; Hollander, V.P. Acid phosphatase from rat liver: Purification, crystallization, and properties. *J. Biol. Chem.* **1968**, *243*, 6084–6089.
49. Gazizova, N.I.; Petrova, N.V.; Karimova, F.G. Effect of tungstate on pea root growth and protein tyrosine phosphorylation. *Russ. J. Plant Physiol.* **2013**, *60*, 776–784, doi:10.1134/s1021443713050051.
50. Kaiser, K.; Wemheuer, B.; Korolkow, V.; Wemheuer, F.; Nacke, H.; Schöning, I.; Schrupf, M.; Daniel, R. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci. Rep.* **2016**, *6*, 33696, doi:10.1038/srep33696.
51. Villamizar, G.A.C.; Nacke, H.; Daniel, R. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity. In *Metagenomics: Methods and Protocols*; Streit, W.R., Daniel, R., Eds.; Springer: New York, NY, USA, 2017; Volume 1539, pp. 249–260.
52. Rutherford, K.; Parkhill, J.; Crook, J.; Horsnell, T.; Rice, P.; Rajandream, M.-A.; Barrell, B. Artemis: Sequence visualization and annotation. *Bioinformatics* **2000**, *16*, 944–945, doi:10.1093/bioinformatics/16.10.944.
53. Wheeler, D.L.; Church, D.M.; Federhen, S.; Lash, A.E.; Madden, T.L.; Pontius, J.U.; Schuler, G.D.; Schriml, L.M.; Sequeira, E.; Tatusova, T.A.; et al. Database resources of the national center for biotechnology. *Nucleic Acids Res.* **2003**, *31*, 28–33, doi:10.1093/nar/gkv1290.
54. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. Interproscan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240, doi:10.1093/bioinformatics/btu031.
55. Marchler-Bauer, A.; Zheng, C.; Chitsaz, F.; Derbyshire, M.K.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Lanczycki, C.J.; et al. CDD: Conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **2013**, *41*, D348–D352, doi:10.1093/nar/gks1243.
56. Wolstencroft, K.J.; Stevens, R.; Taberner, L.; Brass, A. Phosphabase: An ontology-driven database resource for protein phosphatases. *Proteins* **2005**, *58*, 290–294, doi:10.1002/prot.20325.
57. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. Signalp 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786, doi:10.1038/nmeth.1701.
58. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nat. Commun.* **2016**, *7*, 11257, doi:10.1038/ncomms11257.
59. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezuk, Y.; et al. Blast: A more efficient report with usability improvements. *Nucleic Acids Res.* **2013**, *41*, W29–W33, doi:10.1093/nar/gkt282.

60. Mitchell, A.L.; Scheremetjew, M.; Denise, H.; Potter, S.; Tarkowska, A.; Qureshi, M.; Salazar, G.A.; Pesseat, S.; Boland, M.A.; Hunter, F.M.I.; et al. EBI metagenomics in 2017: Enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res.* **2017**, *46*, D726–D735
61. Edgar, R.C. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
62. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. Mega X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549, doi:10.1093/molbev/msy096.
63. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245, doi:10.1093/nar/gkw290.
64. Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-Tasser suite: Protein structure and function prediction. *Nat. Methods* **2015**, *12*, 7–8, doi:10.1038/nmeth.3213.
65. Zhang, Y. I-Tasser server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40, doi:10.1186/1471-2105-9-40.
66. Cold Spring Harbor Protocols. *LB (Luria-Bertani) Liquid Medium*; Cold Spring Harbor Protocols: Cold Spring Harbor, NY, USA, 2006; doi:10.1101/pdb.rec8141.
67. Cold Spring Harbor Protocols. *M9 Minimal Medium (Standard)*; Cold Spring Harbor Protocols: Cold Spring Harbor, NY, USA, 2010; doi:10.1101/pdb.rec12295.
68. Heinonen, J.K.; Lahti, R.J. A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase. *Anal. Biochem.* **1981**, *113*, 313–317, doi:10.1016/0003-2697(81)90082-8.
69. Greiner, R. Purification and properties of a phytate-degrading enzyme from *Pantoea agglomerans*. *Protein J.* **2004**, *23*, 567–576, doi:10.1007/s10930-004-7883-1
70. Gomori, G. Preparation of buffers for use in enzyme studies. In *Methods in Enzymology*; Academic Press: Cambridge, MA, USA, 1955; Volume 1, pp. 138–146
71. Taberero, L.; Aricescu, A.R.; Jones, E.Y.; Szedlacsek, S.E. Protein tyrosine phosphatases: Structure-function relationships. *FEBS J.* **2008**, *275*, 867–882, doi:10.1111/j.1742-4658.2008.06251.x



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Supplemental Information for Chapter 4*Contents*

Table S1. Protein sequence similarities of deduced gene products encoded by the insert of pLP16

Table S2. Supplementary Table S2. Rank of templates representing the top ten threading PDB templates used by I-TASSER.*

Fig. S1. A) Best 3D structure model of the Pho16B protein, predicted by I-TASSER and edited with Pymol 2.2, calculated C score -1.22. B) Amino acids involved in the MptpB-loop in Pho16B.

Table S1. Protein sequence similarities of deduced gene products encoded by the insert of pLP16

ORF number	No. of encoded amino acids	Accession no, closest similar protein, organism, (no. of encoded amino acids),	Percent identity to the closest similar protein (BlastP)
<i>pho16B</i>	376	WP_042381880.1,protein-tyrosine phosphatase <i>Streptacidiphilus melanogenes</i> , (372)	(79%)
<i>Orf02</i>	325	WP_042417920.1, LacI family transcription/Al regulator <i>Streptacidiphilus anmyonensis</i> , (325)	(82%)
<i>Orf03</i>	117	WP_018557579.1, hypothetical protein <i>Streptomyces sp.</i> BoleA5 (883)	(53%)
<i>Orf04</i>	119	RJW72075, vacuolar protein sorting-associated protein 53 -like protein <i>Clonorchis sinensis</i> ,(1358)	(29%)
<i>Orf05</i>	37	PZN39371.1, sulfoN/Ate ABC transporter ATP-binding protein, Proteobacteria bacterium, (246)	(63%)
<i>Orf06</i>	96	WP_099477000.1, TerC family protein <i>Paenibacillus ihbetae</i> , (221)	(34%)
<i>Orf07</i>	38	AAW27310.1, SJCHGC02811 protein <i>Schistosoma japonicum</i> , (133)	(45%)
<i>Orf08</i>	102	No BLAST hits found	N/A
<i>Orf09</i>	130	XP_016657880.1, uncharacterized protein LOC100159331 isoform X3 <i>Acyrtosiphon pisum</i> , (6853)	(42%)
<i>Orf10</i>	50	No BLAST hits found	N/A
<i>Orf11</i>	818	AWN00235.1, hypothetical protein, uncultured organism, (819)	(99%)
<i>Orf12</i>	35	WP_109505691.1, class I SAM-dependent methyltransferase <i>Nocardioides sp. YIM ART13</i> , (270)	(65%)
<i>Orf13</i>	200	No BLAST hits found	N/A
<i>Orf14</i>	41	No BLAST hits found	N/A
<i>Orf15</i>	109	No BLAST hits found	N/A
<i>Orf16</i>	102	No BLAST hits found	N/A
<i>Orf17</i>	63	OHB94183.1, transcription termination factor NusA <i>Planctomycetes bacterium</i> , (358)	(28%)
<i>Orf18</i>	191	XP_001794145.1, hypothetical protein SNOG_03588 <i>Parastagonospora nodorum SN15</i> (641)	(36%)
<i>Orf19</i>	118	PQE15591.1, Heavy metal tolerance protein <i>Rutstroemia sp. NJR-2017a BBW</i> , (1031)	(47%)
<i>Orf20</i>	25	KMP12518.1, hypothetical protein UZ36_00105 <i>Nitrospina sp.</i> SCGC, (342)	(100%)
<i>Orf21</i>	40	No BLAST hits found	N/A
<i>Orf22</i>	45	No BLAST hits found	N/A
<i>Orf23</i>	410	WP_090595989.1, methyltransferase domain-containing protein <i>Auraticoccus monumenti</i> , (274)	(54%)
<i>Orf24</i>	140	No BLAST hits found	N/A
<i>Orf25</i>	82	No BLAST hits found	N/A
<i>Orf26</i>	133	No BLAST hits found	N/A
<i>Orf27</i>	48	No BLAST hits found	N/A
<i>Orf28</i>	239	ONH24413.1, hypothetical protein BL253_30125 <i>Frankia asymbiotica</i> , (395)	(80%)
<i>Orf29</i>	64	WP_098051725.1, LacI family transcription regulator <i>Pantoea agglomerans</i> , (341)	(39%)
<i>Orf30</i>	105	No BLAST hits found	N/A
<i>Orf31</i>	117	No BLAST hits found	N/A
<i>Orf32</i>	74	No BLAST hits found	N/A
<i>Orf33</i>	49	No BLAST hits found	N/A
<i>Orf34</i>	69	No BLAST hits found	N/A
<i>Orf35</i>	63	No BLAST hits found	N/A
<i>Orf36</i>	110	KQB60501.1, hypothetical protein AE621_04680 <i>Acidovorax sp</i> , SD340 (145)	(42%)
<i>Orf37</i>	186	No BLAST hits found	N/A
<i>Orf38</i>	74	WP_125089351.1, hypothetical protein <i>Saccharopolyspora sp H219</i> , (134)	(52%)
<i>Orf39</i>	26	PYO22201.1, DNA-3-methyladenine glycosylase I , <i>Candidatus Rokubacteria</i> , 194	(100%)
<i>Orf40</i>	186	No BLAST hits found	N/A

N/A: not applicable

Table S2. Supplementary Table S2. Rank of templates representing the top ten threading PDB templates used by I-TASSER.*

Rank	PDB Hit	Iden1	Iden2	Coverage	Norm. Z-score
1	2oz5	0.33	0.25	0.62	1.43
2	1ywf	0.32	0.23	0.60	3.46
3	5vgr	0.09	0.18	0.93	1.36
4	2oz5	0.33	0.25	0.62	1.87
5	4rkk	0.17	0.16	0.64	1.89
6	1ywf	0.33	0.23	0.60	4.96
7	1ywf	0.33	0.23	0.61	4.15
8	5gnr	0.10	0.17	0.87	1.34
9	2oz5	0.32	0.25	0.62	2.06
10	4rkk	0.14	0.16	0.83	1.12

Iden1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence. Iden2 is the percentage sequence identity of the whole template chains with query sequence. Coverage of the threading alignment is equal to the number of aligned residues divided by the length of query protein. Norm. Z-score is the normalized Z-score of the threading alignments. The top 10 alignments reported above (in order of their ranking) are from the following threading programs: 1: MUSTER, 2: FFAS-3D, 3: PROSPECT2, 4: SPARKS-X, 5: HHSEARCH, 6: HHSEARCH2, 7: HHSEARCH I, 8: PROSPECT2, 9: Neff-PPAS, 10: SP3.

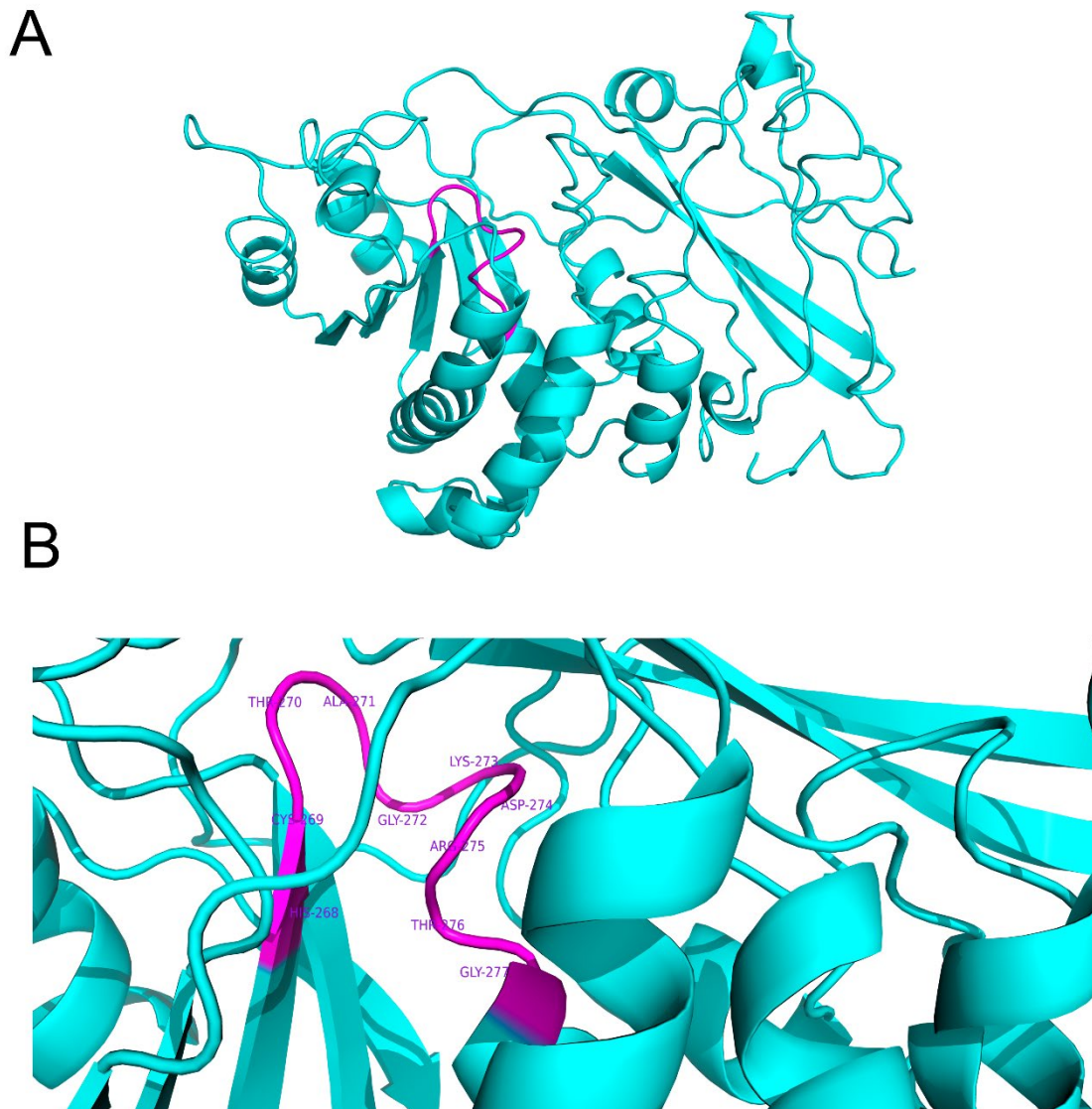


Fig. S1. A) Best 3D structure model of the Pho16B protein, predicted by I-TASSER and edited with Pymol 2.2, calculated C score -1.22. B) Amino acids involved in the MptpB-loop in Pho16B.

5**Functional Metagenomics Unwraps a New Catalytic Domain Associated to Phytase Activity: the Metallo- β -lactamase Superfamily Domain**

Genis Andrés Castillo Villamizar^{1,2,§}, Katrina Funkner^{1,§}, Heiko Nacke¹, Karolin Foerster and Rolf Daniel¹

Accepted mSphere

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany,

²Línea tecnológica biocorrosión, Corporación para la investigación de la corrosión C.I.C. Piedecuesta, Santander, Colombia.

§Both authors contributed equally to this work

Author contributions to the work:

Performed the experiments: KF, GAC, KFo.

Analyzed data: GAC, KF, HN.

Wrote the paper: GAC, KF, HN, RD

Conceived the concept and designed the experiments: HN, RD.

5.1. Abstract

Phytic acid is an alternative phosphorus source that plays several important roles in the phosphorus cycle and in cell metabolism. The known characterized enzymes responsible for its processing, the phytases, are mostly derived from cultured individual microorganisms. The catalytic signatures of phytases are restricted to the molecular domains of four protein superfamilies: histidine phosphatases, protein tyrosine phosphatases, the purple acid phosphatases and the β -propeller phosphatases. During function-based screening of previously generated soil metagenomic libraries for *E. coli* clones conferring phytase activity two library-bearing clones harbouring the plasmids pLP05 and pLP12 were detected. Analysis of the insert sequences revealed the absence of classic phosphatase/phytase signatures of the proteins deduced from the putative genes but the genes *mblp01* (pLP05) and *mblp02* (pLP12) encoded putative metallo- β -lactamases (MBLs). Several MBL representatives are promiscuous proteins with phosphodiesterase activity but phytase activity was previously not reported. Both *mblp01* and *mblp02* were subcloned and expressed and the gene products were analysed. Mblp01 and Mblp02 are members of the lactamase B2 family. The closest structural homologue of both proteins was the ZipD protein of *Escherichia coli*. Mblp01 and Mblp02 showed activity towards almost all analysed substrates including phytate. The maximal enzyme activities were recorded for Mblp01 at 50°C under acidic conditions, and for Mblp02 at 35°C and a neutral pH value. In the presence of Cu^{2+} or SDS the enzymatic activities of Mblp01 and Mblp02 were strongly inhibited. Analyses of the minimal inhibitory concentration of several β -lactam antibiotics revealed that recombinant *E. coli* cells carrying *mblp01* or *mblp02* showed reduced sensitivity towards β -lactam antibiotics.

5.2. Importance

Phytic acid is a main phosphorus storage molecule in many plant tissues, an alternative source of phosphorus and a problematic anti-nutrient. In addition, it influences functions of secondary messengers and acts as antioxidant in tumor growth prevention. The enzymatic capability to process phytate has been reported for a limited number of protein families. This might be due to the almost exclusive use of proteins derived from individual microorganisms for the analysis of phytase activity. With such a restriction, the study of the complexity and diversity of the phytases remains incomplete. By using metagenome-

derived samples, this study demonstrates the existence of phytase activity in one of the most promiscuous superfamilies, the metallo- β -lactamases. Our results increase the general knowledge on phytase diversity in environmental samples and could provide new avenues for the study and engineering of new biocatalysts.

KEYWORDS phytate, β -lactamase, phosphatase, phytase, metallo- β -lactamases, soil metagenome

5.3. Introduction

Over the last two centuries, anthropogenic activities have altered the global biogeochemical cycles of elements such as carbon (C), nitrogen (N) and phosphorus (P) significantly. With the accelerated global population growing rates, the P flow to the biosphere has been quadrupled. This cycle disruption is mainly due to the mining of P compounds for fertilizers. P sources such as phosphorus-rocks are not renewable and the current resources are being rapidly depleted (1, 2). Consequently, a better understanding of the P cycle, and development of strategies and alternatives for P acquisition are of increasing importance.

Phytic acid, also known as inositol-6-phosphate or phytate in its salt form represents an immense and almost unexploited reservoir of P that could be potentially utilized by plants, microorganisms and animals (3). In humans, phytate and some of its degradation products have been related to metabolic effects such as the prevention of kidney stone formation and a possible protection against diseases such as diabetes mellitus, atherosclerosis, coronary heart disease and some types of cancer (4, 5). Phytate is the most abundant source of P in several types of cereals and grains that are commonly used to feed animals employed for large-scale production. However, monogastric animals such as swine, poultry, and fish cannot digest phytate efficiently due to the lack of phytases. This limited capacity of phytate processing has led to severe environmental problems, i.e. P eutrophication of water bodies (3, 6).

Since the cleavage of the phosphate residues from phytate requires phytases, the search for new phytases with novel and improved characteristics has been one focus of phosphatase research. Several phytases have been reported and characterized, revealing the existence of different catalytic mechanisms to cleave phosphate groups from phytate. Some phytases are used commercially to reduce the impact of phytate accumulation.

Nevertheless, the *in vivo* functions of phytate and several phytate hydrolysis mechanisms are not fully known (7).

Phytase activity has been reported to be restricted to only four classes of protein phosphatases with different catalytic mechanisms. The histidine acid phosphatases (HAP-phy) represent the most extensively studied class of phosphatases to which all commercially used phytases belong (8). Another group of phytases comprises β -propeller phosphatases (BPP-phy), which are considered as a relatively new class of enzymes and exhibit no significant homology to any known phosphatases. Furthermore, representatives of purple acid phosphatases (PAP-phy) that are mostly found in plants, and the protein tyrosine phosphatases (PTPs-phy or cysteine phytases) that are the main phytate-degrading enzymes of ruminant animals (9), are known to exhibit phytase activity. In addition, several phytases have been associated to microbial pathogenicity in different species (10-13). The vast majority of reported phytases are derived from a small culturable fraction of microorganisms. Recent reports implementing functional metagenomic approaches have evidenced the potential of environmental samples as source of novel phosphatases/phytases (14, 15).

In this study, two environmentally derived promiscuous enzymes (Mblp01 and Mblp02) exhibiting phytate-degrading activity and carrying the catalytic domain of metallo- β -lactamases (MBLs) were characterized. To our knowledge, this superfamily domain has never been associated with this type of activity before. MBLs catalyze the hydrolysis of a wide range of substrates such as β -lactam antibiotics including penicillin and also the latest generation of cephalosporins. Members of this family also comprise glycosylase, lactonase, arylsulfatase, 5'-exonuclease or ribonuclease activities. In addition, phosphoesterase activity has been reported (16).

5.4. Results

Identification and sequence analysis of the novel phytase-encoding genes

Two phosphatase/phytase active *E. coli* clones harbouring the recombinant plasmids pLP05 (2,496 bp) and pLP12 (5,578 bp) were recovered by functional screening of soil metagenomic libraries using phytic acid as sole P source (14). Sequence analysis of the inserts did not reveal putative genes similar to known ones encoding phosphatase/phytase activity, which could be responsible for the activity of the

recombinant *E. coli* strains on indicator agar with phytic acid as P source (Fig. S1; Fig. S2). However, the putative genes *mblp01* (pLP05) and *mblp02* (pLP12) encode proteins carrying the MBL fold. Considering the high promiscuity level of MBLs and the phosphoesterase activity, previously reported for some members of this superfamily (16, 17), both genes were individually subcloned and the thereby generated recombinant *E. coli* strains screened for activity on indicator agar. *E. coli* clones carrying the plasmid with the subcloned *mblp01* or *mblp02* gene showed activity on indicator agar with phytic acid as sole P source, demonstrating that *mblp01* and *mblp02* encode proteins with phosphatase/phytase activity.

The taxonomic classification of the complete inserts indicated that both inserts are of bacterial origin. The insert sequence of pLP05 is affiliated to the phylum *Proteobacteria* and that of pLP12 to *Acidobacteria* (Fig. S3). The genes encoding the predicted MBL fold *mblp01* (pLP05) and *mblp02* (pLP12) encode polypeptides of 312 and 355 amino acids with estimated molecular masses of 33 and 38 kDa, respectively. The protein sequence analysis against the InterPro database showed that gene products carry the metallo- β -lactamase domain (IPR001279). Neither of the two proteins could be assigned to any specific family during this analysis, but the signature of the homologous superfamily ribonuclease Z/hydroxyacylglutathione hydrolase-like (IPR036866) was present in both sequences. The search against the Pfam database assigned both proteins to the lactamase B2 family (PF12706). The domain organization for both proteins derived from the Pfam analysis is shown in Fig. 1. The sequence similarity searches showed that Mblp01 and Mblp02 exhibited 61% amino acid sequence identity to a hypothetical protein associated to the phylum *Verrucomicrobia* (PYI90218.1) and 51% to a metallo-hydrolase from *Blastomonas* sp. (WP_054133775.1), respectively. In addition, we performed a search against metagenome databases in order to find the closest homologous of Mblp01 and Mblp02 in environmental samples. Screening of the metagenome-derived protein database from the NCBI (env_nr) yielded sequence identities of 39% (Mblp01) and 49% (Mblp02) to hypothetical proteins derived from marine metagenomes (data not shown). The search against the metagenomic EMBL-EBI database showed a hit (MGYS00000776) with 71.6% sequence identity to Mblp01. MGYS00000776 is derived from a soil metagenome associated to a Brazilian soil forest. In the case of Mblp02, the best hit (49% sequence identity) was to the sequence MGYP000565410107, which originated from a geothermal spring water metagenome from India.

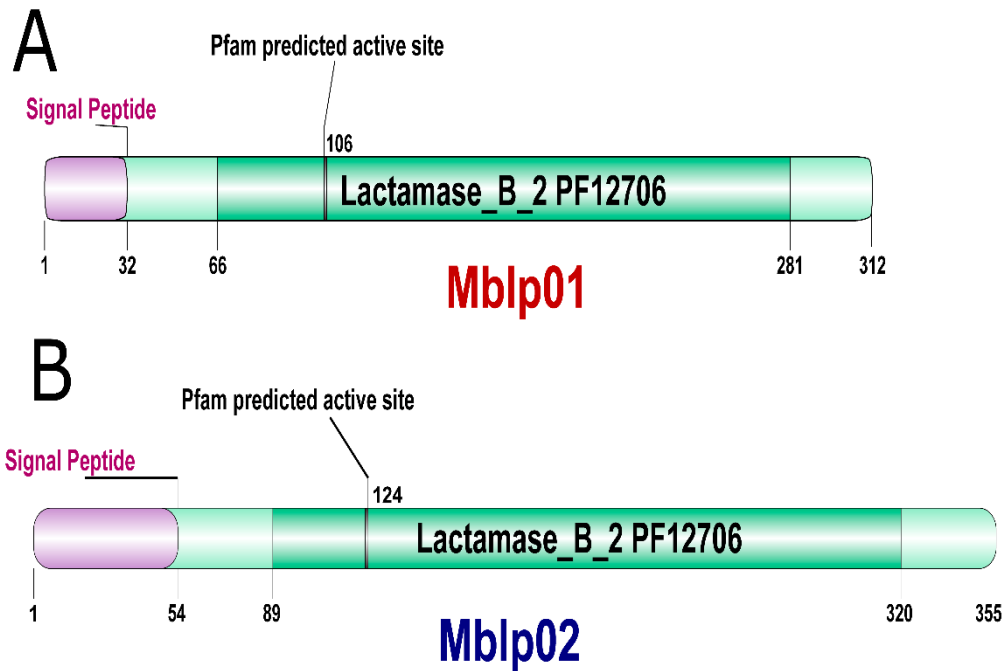


Fig. 1. Domain organization of A) Mblp01 and B) Mblp02.

Phylogenetic positioning of the new MBL representatives Mblp01 and Mblp02

To elucidate the relationship between Mblp01 and Mblp02 and other lactamases of the B2 family and classic phytase representatives, we performed a phylogenetic analysis. Sequences of the Pfam seed group that belong to the B2 sub-class (PF21706), sequences of the B2 *cphA* metallo β -lactamase of *Aeromonas sp.* (18) and representatives of the classical phytases were included in the analysis. Mblp01 and Mblp02 grouped separately in the generated phylogenetic tree (Fig. 2) (Supplementary Extended Data Table 1). Mblp01 formed a small cluster with a B2 family MBL of *Granulicella mallensis* (G8NYQ4). Mblp02 does not group closely with any of the analysed proteins, but it is part of a well-supported large cluster that includes phosphatases/phytases and the *cphA* β -lactamases.

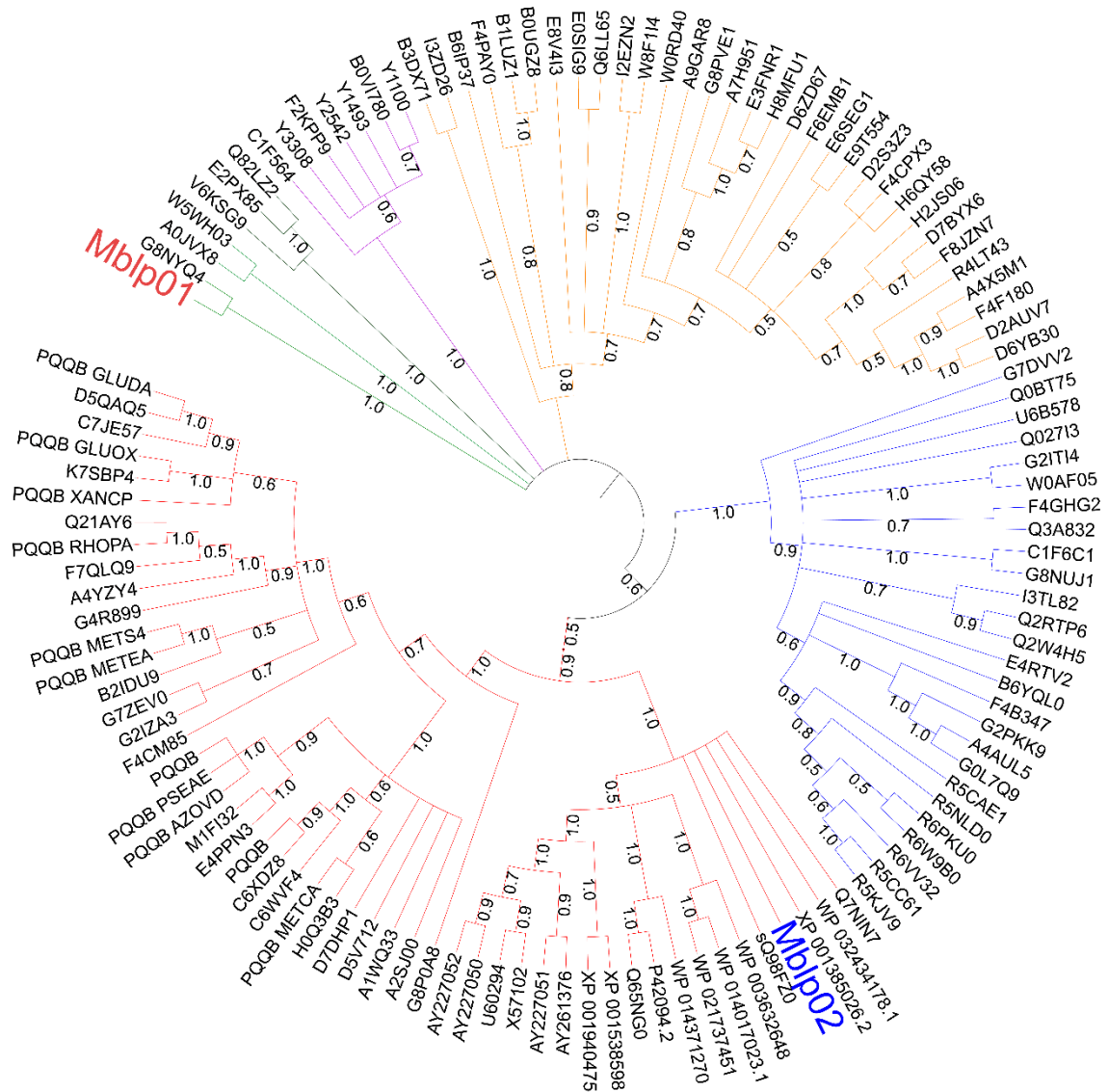


Fig. 2. Neighbor-joining phylogenetic tree showing the position of Mblp01 and Mblp02. The numbers at the nodes indicate levels of bootstrap support (range from 0 to 1) and were based on 500 replicates. Names refer to the corresponding UniProtKB codes of the supplementary extended data Table 1.

Mblp01 and Mblp02 share the same structural analogue: a ZipD protein. To determine the closest structural relatives, we predicted 3D models of Mblp01 and Mblp02. This analysis was performed by using the I-TASSER software suite (19, 20). The best predicted models of the candidates Mblp01 and Mblp02 showed C-score values of -0.75 and -1.70, respectively (Fig. S4). For both proteins the predicted models were linked to the same Protein Data Bank (PDB) entry (2CBN) (21). This entry corresponds to the crystal structure of a zinc phosphodiesterase (ZipD) from *E. coli*, which is involved in the tRNA maturation process. As the native structures of Mblp01 and Mblp02 are not known, the quality of the modelling prediction was determined by calculating the distance between

the predicted models and published native structures. In our case, I-TASSER predicted the quality of the model by calculating the TM-score. TM-scores of 0.783 and 0.697 were calculated for Mblp01 and Mblp02, respectively, indicating a similar structure of the candidate proteins and the reference protein 2CBN.

Enzymatic properties of the first reported MBL representatives showing phytase activity

Mblp01 and Mblp02 were purified by using a combination of affinity chromatography and size exclusion ultrafiltration. The maximum activity of purified Mblp01 and Mblp02 purified with phytic acid under standard reaction conditions was 50 and 35 °C, respectively (Fig. 3). Mblp01 did not show activity at temperatures below 10 and above 60 °C. The activity of Mblp02 increased continually from 10 to 35 °C. At higher temperatures, the activity of the enzyme decreased rapidly and was not detectable at 65 °C. Dependence of phytase activity on pH was determined in the pH range from 2 to 9 at the optimal temperature of each enzyme activity (Fig. 4). Mblp01 showed activity from pH 2 to 7 and Mblp02 from 3.6 to 8. In the case of Mblp01, more than 70% activity was retained between pH 4 and 6 with a maximum activity at pH 5. Mblp02 showed a single narrow activity peak at pH 7.0. Enzymatic activity at pH lower or higher than 7.0 dropped under 60 and 40%, respectively. The ability of Mblp01 and Mblp02 to hydrolyse different phosphorylated compounds was determined under the respective optimal pH and temperature of enzyme activity. Mblp01 showed activity with all tested substrates whereas Mblp02 with all substrates, except pyridoxal phosphate. For both enzymes, the substrates ATP and glucose 6-phosphate yielded the highest activities (Fig. 5). With phytate as substrate, Mblp01 and Mblp02 showed 18 ± 3.1 and 11 ± 1.2 % relative activity, respectively (Fig. 5). The kinetic parameters of both proteins were determined using the purified protein and phytic acid as substrate. K_m values of Mblp01 and Mblp02 were 1.63 ± 0.031 and 0.4 ± 0.03 mM, respectively. The catalytic efficiencies k_{cat}/K_m ($\text{min}^{-1} \text{M}^{-1}$) were 159 ± 12 and 367 ± 19 $\text{mM}^{-1}\text{min}^{-1}$, respectively.

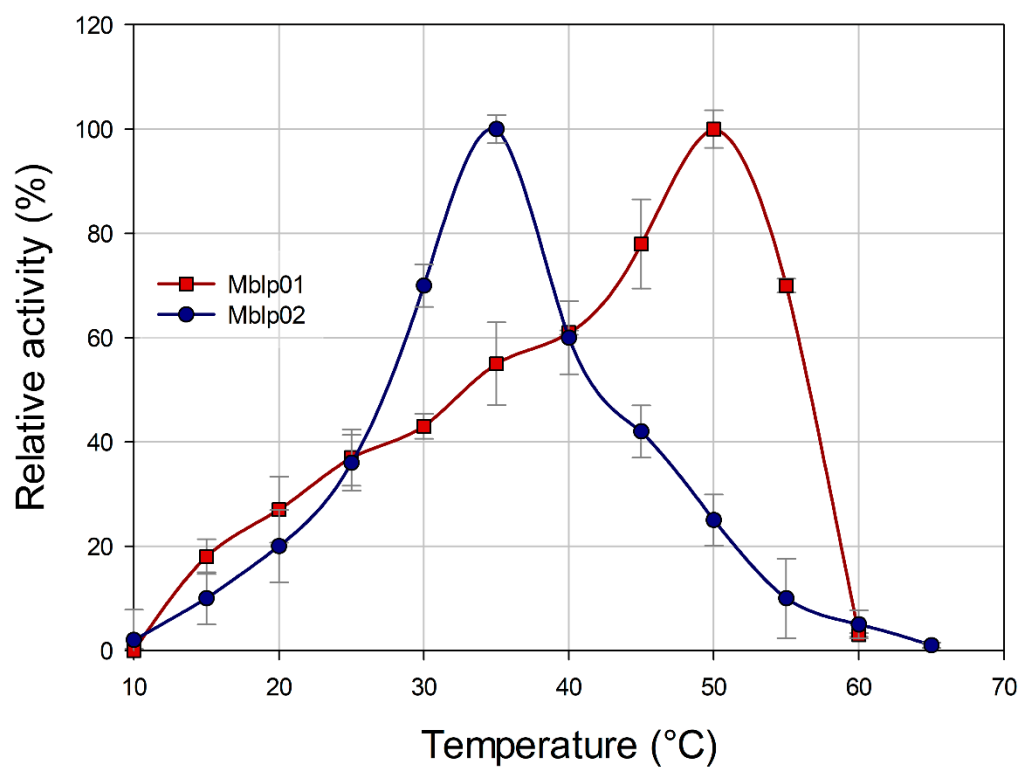


Fig. 3. Effect of temperature on the activity of Mblp01 and Mblp02. All measurements were performed in triplicate using the phytase standard assay at temperatures between 10 and 65°C. 100% relative activity represented 1.92 ± 0.034 and 1.51 ± 0.069 U/mg for Mblp01 and Mblp02, respectively.

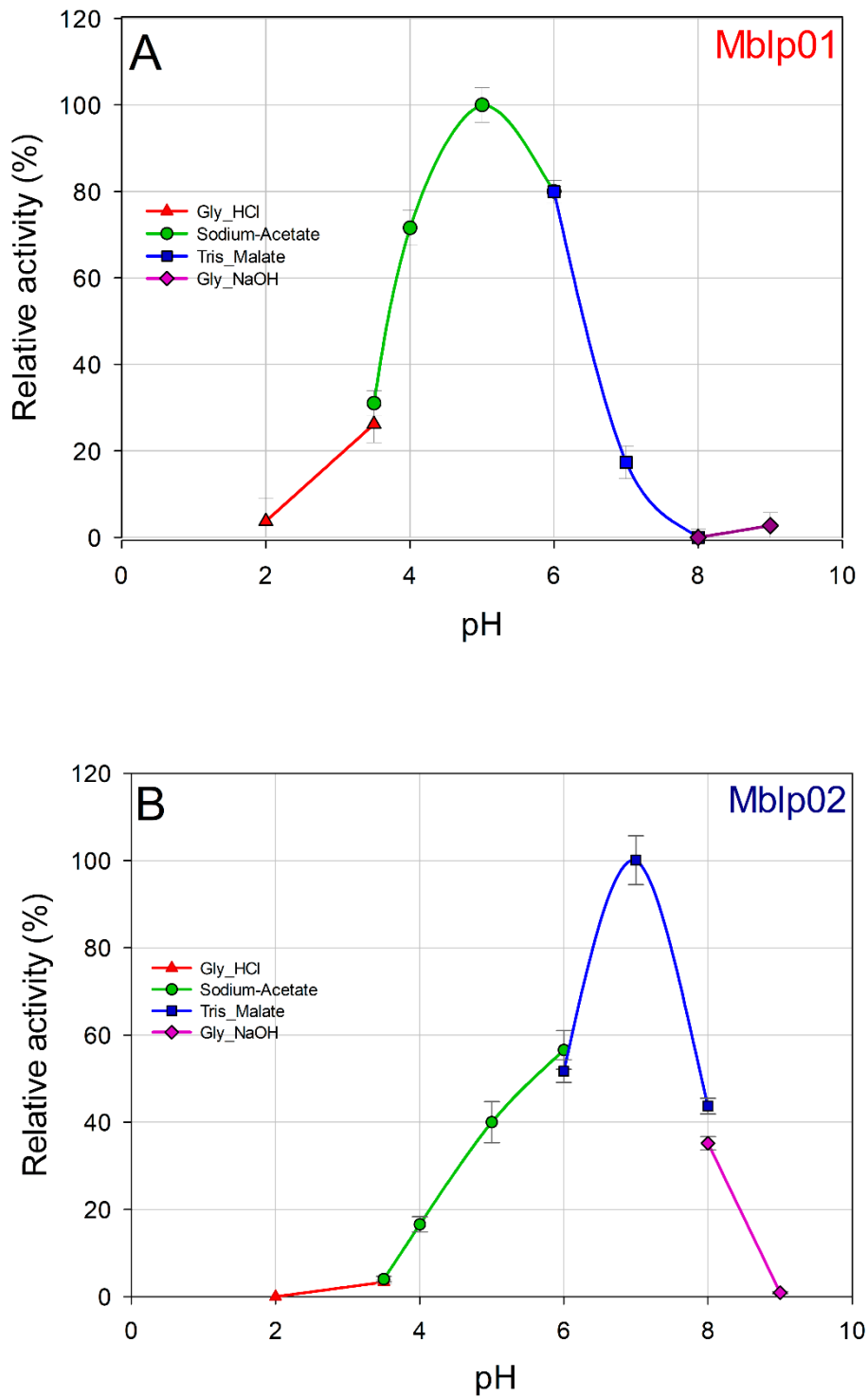


Fig. 4. Effect of pH on activity of (A) Mblp01 and (B) Mblp02. The measurements were performed with different buffer systems according to the phytase standard assay at the optimal temperature of each protein. The average from triplicate experiments with the mean deviation is presented. 100% relative phytase activity represented 1.78 ± 0.016 and 2.1 ± 0.031 U/mg for Mblp01 and Mblp02, respectively.

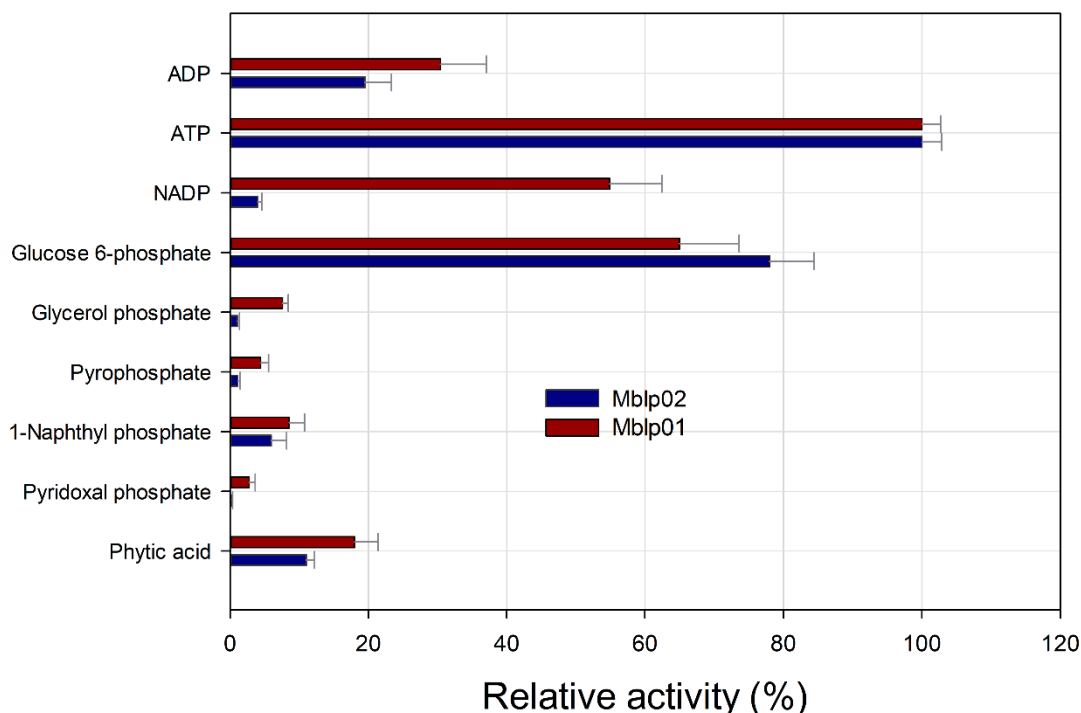


Fig. 5. Substrate specificity of Mblp01 and Mblp02. Specific activities corresponding to 100% of activity with ATP as substrate were 1.77 ± 0.019 and 2.32 ± 0.021 U/mg, respectively. All measurements were performed in triplicate and under optimal pH and temperature conditions for each enzyme.

The effects of various additives on Mblp01 and Mblp02 enzyme activities are summarized in Fig. S5 and Table S1. None of the tested additives enhanced the activity of Mblp01 and Mblp02. Of the evaluated cations, only concentrations higher than 0.5 mM of Cu^{2+} and Fe^{2+} showed a significant inhibitory effect on the activity of both enzymes. The presence of SDS and DTT reduced the enzyme activity of both enzymes below 15%.

Resistance towards β -lactam antibiotics

Taking into account the promiscuous characteristics of MBLs, the phylogenetic positioning of Mblp01 and Mblp02 and the evolutionary relationship between phosphatases and enzymes that can degrade β -lactam antibiotics (22, 23), the ability of *mblp01* and *mblp02* to confer resistance towards β -lactam antibiotics was tested. For this purpose, minimal inhibitory concentrations (MIC) and inhibition zones (halos) in the presence of different β -lactam antibiotics were determined. To obtain an indication whether *mblp01* and *mblp02* confer any level of antibiotic resistance or not, the host strains *E. coli* and *Shimwellia blattae* (formerly *E. blattae*) (24) harbouring the above-generated recombinant plasmids pBAD202-*mblp01* and pBAD202-*mblp02* were evaluated. Strains carrying the cloning vector without insert were used as controls (Table

1). *E. coli* and *S. blattae* strains harbouring *mblp01* or *mblp02* were less sensitive towards most of the evaluated antibiotics such as ampicillin and amoxicillin than the control (Table 1). Thus, a β -lactamase activity was indicated under the tested conditions and suggested that MblP01 and MblP02 are promiscuous enzymes.

Table 1. Sensitivity of *E. coli* and *S. blattae* strains harbouring the *mblp01*- and *mblp02*-containing plasmids (pBAD202-*mblp01* or pBAD202-*mblp02*) and as control the cloning vector without insert (pBAD202) against β -lactam antibiotics.

Strain/construct	MIC ($\mu\text{g/ml}$) ^a				HIZ (mm) ^b	
	Ampicillin	Imipenem	Ceftazidime	Amoxycillin	Doripenem*	Fosfomycin**
<i>E. coli</i> / pBAD202	≥ 1	≥ 1	≥ 0.2	≥ 8	15	12
<i>E. coli</i> / pBAD202- <i>mblp01</i>	≥ 8	≥ 1	≥ 1	≤ 16	9	9
<i>E. coli</i> / pBAD202- <i>mblp02</i>	≥ 8	≤ 4	≥ 8	≤ 16	9	7
<i>S. blattae</i> / pBAD202	≥ 2	≥ 0.5	≥ 0.2	≥ 4	13	10
<i>S. blattae</i> / pBAD202- <i>mblp01</i>	≥ 4	≥ 0.5	≥ 1	≥ 8	9	8
<i>S. blattae</i> / pBAD202- <i>mblp02</i>	≥ 4	≥ 2	≥ 1	≥ 4	6	5

^aMIC, Minimal inhibitory concentration of selected beta-lactam antibiotics; ^bHIZ, diameter of inhibition zones (halos) in relation to each utilized antibiotic. Antibiotic concentration on test disc, *10 and **50 μg . Values are the average of two experiments.

5.5. Discussion

Certain types of phytases are responsible for phytate breakdown during seed germination to make phosphate and *myo*-inositol available for plant nutrition and development (25). The microbial production of extracellular phytase improves phosphate availability in plant roots and overcome phosphate starvation for example in *Phaseolus vulgaris* (26, 27). In addition, phytases have been reported to be involved in signal transduction, cell division and microbial pathogenesis (11, 12, 28). Despite their broad relevance, phytase research has been mainly focused on improvement of enzymes for use as animal feed additives. Very little is known about the phytase diversity or their role *in vivo* (29). Phosphatases are in general enzymes that hydrolyse a broad spectrum of phosphorylated compounds including phytate. However, reported phytase activity is limited as above-mentioned to a few protein types (7). One limiting factor for finding new types of proteins or catalytic domains associated with phytase activity is the almost exclusive usage of individual microorganisms for the isolation and characterization of this type of enzyme (14, 30).

Our analysis showed that the metagenome-derived enzymes Mblp01 and Mblp02 carry the MBL catalytic domain and originate from bacteria. Both corresponding genes were derived from forest soil metagenomes. Searches of metagenomes revealed that Mblp01 is similar to a protein deduced from a forest soil metagenome in Brazil, indicating that enzymes with similar characteristics might be present in other forest soils.

The sequences of Mblp01 and Mblp02 clustered differently in the constructed phylogenetic tree. Mblp01 is part of a small monophyletic cluster together with an uncharacterized MBL protein (G8NYQ4) derived from a genome sequence of *Granulicella mallensis*, which was originally isolated from tundra soil of northwestern Finland (31). Mblp02 was grouped within a larger cluster that contains β -lactam-degrading enzymes from *Aeromonas salmonicida* (e.g. AY261376). Several representatives of protein phytases such as a β -propeller phytase from *Bacillus subtilis* (Q6NG0), or the acid phosphatase from *Scheffersomyces stipitis* (XP_001385026) clustered in the same group (Fig 3). The evolutionary link between MLBs and phosphatases has been reported, but is not fully clear (23). Chakraborty et al. 2012 described the likelihood of the presence of a Class B2 MBL-like scaffold in a cold active alkaline phosphatase from *Vibrio*. The prediction was verified by the inhibition of the phosphatase activity by imipenem. The results suggested a natural evolution of the alkaline phosphatases to acquire true metallo- β -lactamase activity (22).

The predicted models of Mblp01 and Mblp02 provided first insights into the structure of these proteins (Fig S4). The models are supported by the calculated C and TM-scores and show a remarkable relationship of both enzymes with a zinc phosphodiesterase encoded by the gene *elaC* from *E. coli*. ElaC harbours the MBL domain and possesses phosphodiesterase activity. Additionally, it is affiliated to the tRNase Z-Family and acts as a clamp in tRNA binding (21).

Characterization of phytases has shown that these enzymes exhibit a wide range of features. It has been reported that phytases are usually most active within temperature ranges of 45 to 60 °C (30). However, these reported range might be due to that most searches focussed on thermophilic phytases from individual microorganisms, which can be used in industrial applications. With the identification of metagenome-derived phytases, the temperature range of the optimal phytase activity changed. PhyRC001, a metagenome-derived phytase from red rice, showed optimal activity at 30°C (15). The Mblp02 activity optimum was 35 °C. Similarly, another recently described soil

metagenome-originated phytase (Pho07) showed highest activities at lower temperatures (25 to 30 °C) (14). The reported pH range of phytase activity also varies (2.2 to 8.0) whereby phytases of bacterial origin revealed optimal activities between pH 6.5 and 7.5. A similar pH range was recorded for the optimal activities of Mblp01 and Mblp02, which are according to our analysis of bacterial origin. Nevertheless, other soil metagenome-derived bacterial phytases, i.e. Pho07, exhibited a pH optimum at pH 4.0 (14).

MBLs and phosphatases are both considered as promiscuous enzymes with respect to substrate spectrum or in some cases catalytic mechanism (32, 33). It has been estimated that MBLs can catalyze on average 1.5 reactions. Moreover, directed evolution experiments have shown that with a few mutations, the β -lactamases NDM1 and VIM2 are converted to enzymes with a promiscuous phosphonate monoester hydrolase activity (34). Similar might be the case for the promiscuous Mblp01 and Mblp02 with respect to the additional indicated β -lactamase activity of both enzymes. The tested host strains *E. coli* and *S. blattae* carrying and expressing *mblp01* and *mblp02* show less susceptibility against the tested β -lactam antibiotics than the control strains. It has been previously suggested that the MBL superfamily could have evolved from a common ancestor via promiscuous enzymes with a connected catalytic landscape (16, 34). Some zones of sequence space may overlap between multiple catalytic landscapes, including sequences of enzymes that can catalyze more than one activity (catalytic promiscuity). The connections between different catalytic landscapes amend enzymes to evolve and develop new functions (35). Some of the new enzymatic functions in the MBL superfamily such as phosphodiesterase activity evolved rapidly from the current diversity of enzymes with promiscuous activities. Other MBLs have evolved recently on the basis of the appearance of substrates that were not available a few years ago. A good example are phosphotriesterases, which hydrolyze organophosphate pesticides. This type of enzymes evolved and developed activity toward a substrate that did not exist 60 to 80 years ago (34). The β -lactamase activity was invented several times during evolution in independent ways. It has been hypothesized that promiscuous activities provide an immediate evolutionary advantage against β -lactam antibiotics in an environment in which antibiotics are present (34)

Relatively few organisms use phytate as sole P source. Most reported phosphatases with phytase activity are non-specific phosphatases (29). The K_m values of Mblp01 and Mblp02

are in the same range of other reported environmental phosphatases with phytase activity, e.g. Pho07 and Pho16B (14, 36). The catalytic efficiencies towards phytate of Mblp01 and Mblp02 are significantly lower than catalytic efficiencies of other reported phytases derived from cultured single microorganisms (37-39). The studies with the additives revealed that SDS and DTT exhibited deleterious effects on activity of Mblp01 and Mblp02. In presence of SDS, the activity of both enzymes was strongly reduced or not detectable. SDS has been reported to be a strong inhibitor of phytases (40). Anionic detergents bind to proteins and induce structural changes that affect the protein stability and functionality (41). DTT had also a strong inhibitory effect on the phytase activity of Mblp01 and Mblp02. The same effect was observed for the environmental phytase Pho16B (36). DTT acts as chelator of metal ions essential for enzyme activity (42). The cation Cu^{2+} exhibited an inhibitory effect on both enzymes (Fig S5). It has been reported that copper ions directly inactivate other proteins of the metallo- β -lactamase superfamily and also phytases. However, the mechanism leading to this inactivation remains unknown (43).

In conclusion, the applied function-driven metagenomic approach resulted in identification of two representatives of a new type of phytate-degrading enzymes exhibiting a MBL domain structure. To our knowledge, enzymes merging these two characteristics have not been reported. This study shows that MBLs are potentially involved in a previously not reported process, recovering valuable P from phytate. Our data improve the knowledge on the diversity of phytate-degrading enzymes, which is required to gain insights into the relationships among these enzymes. This knowledge could help in the future to design and engineer superior biocatalysts as well as improve our capabilities to solve problems such as the P scarcity and the proliferation of antibiotic-resistant bacteria.

Materials and Methods

Sampling, metagenomic library construction and clone selection

The genes encoding Mblp01 and Mblp02 originate from two soil samples (SEW46 and HEW30) with pH values of 3.29 and 3.86, respectively (44). Both samples were collected from A horizons of beech forest sites in Germany. SEW46 was collected within the Schorfheide-Chorin biosphere reserve, while the HEW30 sample was collected within

Hainich national park. Collection of the samples was performed as previously described by Kaiser et al (2016) (45). Approximately 2.5 g of soil were used for total DNA extraction by employing the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA). The metagenomic libraries were generated using the method described by Nacke et al. (2011)(44), and constructed and screened by Castillo Villamizar et al. (2019) (14). The metagenomic libraries SEW46 and HEW 30 were composed of 38,122 and 53,460 clones, respectively. Libraries were constructed by using the plasmid pCR-XL-TOPO as vector (Invitrogen GmbH, Karlsruhe, Germany). The library-bearing *E. coli* clones were screened by using modified minimal Sperber minimal medium (16 g/liter agar, 10 g/liter glucose, 500 mg/liter yeast extract, 100 mg/liter CaCl₂, and 250 mg/liter MgSO₄). In order to induce phytase activity, phytate (2.5 g/liter) was used as phosphorus source and 25 µg/ml of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) as indicator. Clones with phosphatase/phytase activity turned from white to dark blue within 48 hours (46, 47).

Sequence data analysis

The insert sequences of plasmids pLP05 and pLP12 derived from the libraries SEW46 and HEW30, respectively, were sequenced and analysed. Initially, the taxonomic classification of the complete DNA inserts of pLP05 and pLP12 was performed by using the software KAIJU (48). Next, open reading frame (ORF) prediction was performed using the ORF finder tool provided by the National Center for Biotechnology Information (NCBI) and the ARTEMIS program (49, 50). The results were verified manually by using criteria such as the presence of a ribosome-binding site, GC frame plot analysis and similarity to known genes.

Amino acid sequences deduced from the *mblp01* and *mblp02* gene products were examined for similarities to known protein families and domains by performing searches against the Pfam, InterPro and NCBI collections. Signal peptide prediction was performed using SIGNALP 4.0 (51-53). Mblp01 and Mblp02 were also analysed by using the Basic Local Alignment Search for proteins (BLASTP). Two NCBI databases; non-redundant sequences (nr) and metagenomic proteins (env_nr) were employed. An additional search was performed against the metagenomic platform of the European Institute of Biotechnology (EMBL-EBI) (52, 54). Multiple sequence alignments of Mblp01 and Mblp02 and related MBLs were performed using MUSCLE (55). Evolutionary analyses were conducted in MEGA 7 using the neighbor-joining method (56, 57). The bootstrap consensus tree was inferred from 500 replicates. The evolutionary distances were

computed using the number of differences method. The analysis involved 137 amino acid sequences and a total of 670 positions in the final dataset. Branches with bootstrap values below 50 % were collapsed. The tree was visualized using iTOL v3 (58). A prediction of the tertiary structure of the proteins Mblp01 and Mblp02 was performed by employing the I-TASSER platform (20). The quality of models generated using I-TASSER is based on two major criteria: the confidence score (C-score) and the template modelling score (TM-score) (19). I-TASSER generated five models for each protein. The models were ranked based on the C-score. The C-scores are calculated on the basis of the statistical significance of the threading profile-profile alignment, as well as structure convergence of the assembly simulations. The C-score ranged from -5 to 2. A high C-score value indicates a model with higher confidence (19). The TM-score addresses the structural similarity of two protein models by measuring the global fold similarity. TM-score is less sensitive to local structural variations and its magnitude for random structure pairs is length-independent. The TM-score has a value range of 0 to 1, whereby 1 indicates a perfect match between two structures (19, 59). By calculating the TM-score the structural similarity between the predicted models of Mblp01 and Mblp02 and other published determined structures is estimated. Values close to 0.5 indicate a model of correct topology. In this study, the models with the highest C-score were selected as the best predicted optimized 3D modelling structure.

Protein expression and purification

To facilitate expression and purification, *mblp01* and *mblp02* were cloned into plasmid pBAD202/D-TOPO according to the instructions of the manufacturer (Thermo Fisher Scientific GmbH, Schwerte, Germany). In this way, sequences encoding the His₆ and thioredoxin tags were added to the N terminus of the produced proteins during cloning. As control, a non-coding DNA region was also cloned in pBAD202/D-TOPO. The fidelity of the constructs was confirmed by Sanger Sequencing. The generated constructs were used to transform *Escherichia coli* LMG194. Transformants were grown on Sperber screening media supplemented with 0,2% arabinose. Only the clones carrying a recombinant plasmid harbouring *mblp01* or *mblp02* showed phosphatase/phytase activity on indicator agar after transformation.

The expression plasmids containing *mblp01* or *mblp02* (pBAD202-*mblp01* or pBAD202-*mblp02*, respectively) were transformed into *Escherichia coli* LMG194. Subsequently, the recombinant *E. coli* strains were grown on Luria-Bertani (LB) agar plates supplemented

with kanamycin (50 µg/µl) and incubated at 37°C. One single colony of each construct was used to inoculate 1 L of M9 minimal salts medium (60) containing 50 µg/µl kanamycin and 2% glycerol. The culture was incubated using a New Brunswick Innova 44 incubator-shaker (Eppendorf AG, Hamburg, Germany) with shaking (90 rpm) at 37°C. Protein expression was induced at an OD₆₀₀ of 0.6 using L-arabinose (final concentration, 0.2%). Cells were harvested after 5 h of incubation by centrifugation for 30 min at 4 °C and 8,000 rpm (Sorvall® RC6 centrifuge, rotor SLA 3000, Thermo Fisher Scientific). The resulting cell pellets were suspended in 10 ml of 50 mM HEPES buffer containing 250 mM NaCl and 0.5 mM ZnSO₄. Mechanical cell disruption was performed using a French press (1.38 x 10⁸ Pa; Thermo Fisher Scientific). Subsequently, the extract was cleared by centrifugation for 0.5 h at 4 °C and 15,000 rpm (Sorvall® RC6 centrifuge with rotor SS 35, Thermo Fisher Scientific). The crude extract was filtered using filters with a pore size of 0.45 µm and then 0.2 µm filters (Sarstedt, Nümbrecht, Germany). In order to purify the His₆-tagged protein, the Protino® Ni-TED 2000 purification kit was used as recommended by the manufacturer (Macherey and Nagel, Düren, Germany) with modifications. The equilibration of the columns and the washing steps were performed with 50 mM HEPES (pH 8.0) containing 250 mM NaCl and 0.5 mM ZnSO₄. Elution was performed with 50 mM HEPES containing 250 mM NaCl, 250 mM imidazole and 0.5 mM ZnSO₄. Buffer exchange and imidazole removal was performed by ultrafiltration using Vivaspin® 20 concentrators with exclusion limit of 30 kDa combined with diafiltration cups VS20 as recommended by the manufacturer (Sartorius AG, Göttingen, Germany). The thioredoxin tag of the proteins was removed by employing the enterokinase cleavage capture kit as recommended by the manufacturer (Merck KGaA, Darmstadt, Germany) with a modified cleavage buffer consisting of 50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 2mM CaCl₂ and 0.5 mM ZnSO₄. Subsequently, ultrafiltration of the protein extract using a Vivaspin® concentrator with exclusion limit of 10 kDa was performed.

Biochemical characterization of Mblp01 and Mblp02

Phosphatase activity was determined at 355 nm by detecting the release of inorganic phosphorus according to the ammonium molybdate method developed by Heinonen and Lahti with modifications (61) The purified enzyme solution (10 µl) was pre-incubated for 3 min at 40 °C in 380 µl of 50 mM sodium acetate buffer (pH 5). Subsequently, 10 µl of 100 mM phytic acid dipotassium salt (Sigma-Aldrich, Munich, Germany) was added and the mixture incubated for 30 min at 40 °C. To stop the reaction, 1.5 ml of freshly prepared

AAM solution (acetone–5N H₂SO₄–10 mM ammonium molybdate) and 100 µl 1 M citric acid were added. Samples were measured against blanks prepared by adding AAM solution prior to the addition of enzyme. The absorbance (355 nm) was measured using the Ultrospec® 3300 pro (Amersham plc, Little Chalfont, United Kingdom). All measurements were performed in triplicate. To calculate the enzyme activity, a calibration curve was generated in the range of 5 to 600 nmol phosphate. One activity unit (U) represented the release of 1 nmol phosphate per min.

The influence of temperature on enzymatic activity was determined via the above-described standard phytase assay. The enzymatic activity was evaluated in a temperature range of 10 to 65 °C by using a temperature-adjusted buffer (50 mM sodium acetate, pH 6). In order to analyse the pH dependence of enzyme activity, the following overlapping buffers were prepared as described by Gomori 1955 (62): 50 mM glycine-HCl (pH 2.0, 3.0, and 3.6), sodium acetate (pH 3.6, 4.0, 5.0 and 6.0), Tris-malate (pH 6.0, 7.0, and 8.0), Tris-HCl (pH 8.0, and 9.0), and glycine-NaOH (pH 9.0).

The substrate specificity was determined using the standard phytase assay under the optimal temperature and pH conditions. Nine different substrates comprising ADP, ATP, NADP, glucose-6-phosphate, glycerophosphate, pyridoxal phosphate, pyrophosphate, naphthyl phosphate, and phytic acid were tested using 10 mM concentration. Furthermore, the effect of cations (Al²⁺, Ca²⁺, Cu²⁺, Co²⁺, Fe²⁺ and Mg²⁺) and the potential inhibitors ethylenediaminetetraacetic acid (EDTA), tungstate, oxalate, sodium dodecyl sulfate (SDS), and dithiothreitol (DTT) (1 mM) were analysed. Kinetic parameters K_m and k_{cat}/K_m for both enzymes were calculated from the Michaelis–Menten equation by using the kinetics module of the program SigmaPlot 12.0 (Systat Software, Inc., San Jose, CA, USA). All measurements were performed under optimal pH and temperature conditions using phytate as substrate.

In addition to the biochemical characterization of the enzymes Mblp01 and Mblp02, an analysis of the antimicrobial activity of the gene products encoded by mblp01 and mblp02 were performed. The minimal inhibitory concentration (MIC) for β-lactam antibiotics was performed. Ampicillin, imipenem, ceftazidime and amoxicillin strips (M.I.C.E™) (Oxoid, Wade Road, Basingstoke, U.K) were used. Two additional antibiotics doripenem and fosfomicin (Oxoid, Wade Road, Basingstoke, U.K) were evaluated by measuring the inhibition zone (halos) around discs containing of 10 doripenem or 50 µg fosfomicin. *E. coli* Top 10 and *Shimwellia blattae* DSM 4881 were transformed with the pBAD202-

mblp01 and pBAD202-*mblp02* plasmids. Strains harbouring the cloning vector pBAD202 without insert served as control. All recombinant strains were analysed in duplicate using the M.I.C.Evaluator™ system according to the instructions of the manufacturer (Oxoid, Wade Road, Basingstoke, U.K) with Luria-Bertani (LB) Agar containing 50 µg/ml kanamycin and 0,2% arabinose (pH: 7.0).

Data availability

The sequences of the inserts have been submitted to the National Center for Biotechnology Information (NCBI) under the GenBank accession numbers MH367836 (pLP05) and MH367837 (pLP12).

Acknowledgments

We thank the Deutscher Akademischer Austauschdienst (DAAD) and Colciencias Colombia for the financial support to Genis Andrés Castillo in the frame of the fellowship program ALECOL. We thank Dr. Silja Brady and Mechthild Bömeke for providing technical assistance. We acknowledge support by the Open Access Publication Funds of the University of Göttingen.

5.6. References

1. Falkowski P, Scholes RJ, Boyle E, Canadell J, Canfield D, Elser J, Gruber N, Hibbard K, Högberg P, Linder S, Mackenzie FT, Moore III B, Pedersen T, Rosenthal Y, Seitzinger S, Smetacek V, Steffen W. 2000. The global carbon cycle: A test of our knowledge of earth as a system. *Science* 290:291-296.
2. Smil V. 2000. Phosphorus in the environment: Natural flows and human interferences. *Annu Rev Energ Env* 25:53-88.
3. Riaz A, Chaudhary AN, Hayat R, Hussain Q, Tahir MI, Imran M. 2015. Microbial phytase activity and their role in organic P mineralization. *Arch Agron Soil Sci* 61:751-766.
4. Fox CH, Eberl M. 2002. Phytic acid (IP6), novel broad spectrum anti-neoplastic agent: a systematic review. *Complement Ther Med* 10:229-234.
5. Silva EO, Bracarense APFRL. 2016. Phytic acid: from antinutritional to multiple protection factor of organic systems. *J Food Sci* 81:R1357-R1362.
6. Vats P, Bhattacharyya MS, Banerjee UC. 2005. Use of phytases (*myo*-inositolhexakisphosphate phosphohydrolases) for combatting environmental pollution: a biological approach. *Crit Rev Env Sci Tec* 35:469-486.
7. Mullaney E, Ullah AH. 2007. Phytases: Attributes, catalytic mechanisms and applications, p 97-110. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.
8. Dersjant-Li Y, Awati A, Schulze H, Partridge G. 2015. Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *J Sci Food Agric* 95:878-896.
9. Li Z, Huang H, Zhao H, Meng K, Zhao J, Shi P, Yang P, Luo H, Wang Y, Yao B. 2014. Genetic diversity and expression profiles of cysteine phytases in the sheep rumen during a feeding cycle. *Lett Appl Microbiol* 59:615-620.
10. Bhavsar K, Khire JM. 2014. Current research and future perspectives of phytase bioprocessing. *Rsc Adv* 4:26677-26691.
11. Tsang PW, Fong WP, Samaranayake LP. 2017. *Candida albicans* orf19.3727 encodes phytase activity and is essential for human tissue damage. *PLoS One* 12:e0189219.
12. Blüher D, Laha D, Thieme S, Hofer A, Eschen-Lippold L, Masch A, Balcke G, Pavlovic I, Nagel O, Schonsky A, Hinkelmann R, Worner J, Parvin N, Greiner R, Weber S, Tissier A, Schutkowski M, Lee J, Jessen H, Schaaf G, Bonas U. 2017. A 1-phytase type III effector interferes with plant hormone signaling. *Nat Commun* 8:2159.
13. Ushasree MV, Vidya J, Pandey A. 2017. 14 - other enzymes: phytases, p 309-333. *In* Pandey A, Negi S, Soccol CR (ed), *Current Developments in Biotechnology and Bioengineering*. Elsevier, Amsterdam, Netherlands.
14. Castillo Villamizar GA, Nacke H, Boehning M, Herz K, Daniel R. 2019. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio* 10:e01966-01918.
15. Farias N, Almeida I, Meneses C. 2018. New bacterial phytase through metagenomic prospecting. *Molecules* 23:448.
16. Baier F, Copp JN, Tokuriki N. 2016. Evolution of enzyme superfamilies: comprehensive exploration of sequence–function relationships. *Biochemistry* 55:6375-6388.
17. Vogel A, Schilling O, Niecke M, Bettmer J, Meyer-Klaucke W. 2002. ElaC encodes a novel binuclear Zinc phosphodiesterase. *J Biol Chem* 277:29078-29085.
18. Laconis JP, Sanders CC. 1990. Purification and characterization of inducible beta-lactamases in *Aeromonas* spp. *Antimicrob Agents Chemother* 34:44-51.
19. Yang J, Zhang Y. 2015. Protein structure and function prediction using I-TASSER. *Curr Protoc Bioinformatics* 52:5.8.1-5.8.15.
20. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725-738.

21. Kostecky B, Pohl E, Vogel A, Schilling O, Meyer-Klaucke W. 2006. The crystal structure of the Zinc phosphodiesterase from *Escherichia coli* provides insight into function and cooperativity of tRNase Z-family proteins. *J Bacteriol* 188:1607-1614.
22. Chakraborty S, Ásgeirsson B, Minda R, Salaye L, Frère J-M, Rao BJ. 2012. Inhibition of a cold-active alkaline phosphatase by imipenem revealed by in silico modeling of metallo- β -lactamase active sites. *FEBS Lett* 586:3710-3715.
23. Zhang H, Yang L, Yan L-F, Liao R-Z, Tian W-Q. 2018. Evolution of phosphotriesterase activities of the metallo- β -lactamase family: A theoretical study. *J Inorg Biochem* 184:8-14.
24. Brzuszkiewicz E, Waschkowitz T, Wiezer A, Daniel R. 2012. Complete genome sequence of the B12-producing *Shimwellia blattae* strain DSM 4481, isolated from a cockroach. *J Bacteriol* 194:4436-4436.
25. Azeke MA, Elsanhoty RM, Egielewa SJ, Eigbogbo MU. 2011. The effect of germination on the phytase activity, phytate and total phosphorus contents of some Nigerian-grown grain legumes. *J Sci Food Agric* 91:75-79.
26. Maougal RT, Bargaz A, Sahel C, Amenc L, Djekoun A, Plassard C, Drevon J-J. 2014. Localization of the *Bacillus subtilis* beta-propeller phytase transcripts in nodulated roots of *Phaseolus vulgaris* supplied with phytate. *Planta* 239:901-908.
27. Mukhametzyanova AD, Akhmetova AI, Sharipova MR. 2012. Microorganisms as phytase producers. *Microbiology* 81:267-275.
28. Bruder LM, Gruninger RJ, Cleland CP, Mosimann SC. 2017. Bacterial PhyA protein-tyrosine phosphatase-like *myo*-inositol phosphatases in complex with the Ins(1,3,4,5)P4 and Ins(1,4,5)P3 second messengers. *J Biol Chem* 292:17302-17311.
29. Greiner R. 2007. Phytate-degrading enzymes: Regulation of synthesis in microorganisms and plants, p 78-96. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.
30. Kumar A, Chanderman A, Makolomakwa M, Perumal K, Singh S. 2016. Microbial production of phytases for combating environmental phosphate pollution and other diverse applications. *Crit Rev Env Sci Tec* 46:556-591.
31. Männistö MK, Rawat S, Starovoytov V, Häggblom MM. 2012. *Granulicella arctica* sp. nov., *Granulicella mallensis* sp. nov., *Granulicella tundricola* sp. nov. and *Granulicella sapmiensis* sp. nov., novel acidobacteria from tundra soil. *Int J Syst Evol Microbiol* 62:2097-2106.
32. Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, Toews Keating S, Aono M, Love JD, Evans B, Seidel RD, Hillerich BS, Garforth SJ, Almo SC, Mariano PS, Dunaway-Mariano D, Allen KN, Farelli JD. 2015. Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc Natl Acad Sci U S A* 112:E1974-E1983.
33. Schenk G, Mitić N, Gahan LR, Ollis DL, McGeary RP, Guddat LW. 2012. Binuclear metallohydrolases: complex mechanistic strategies for a simple chemical reaction. *Acc Chem Res* 45:1593-1603.
34. Baier F, Tokuriki N. 2014. Connectivity between catalytic landscapes of the metallo-beta-lactamase superfamily. *J Mol Biol* 426:2442-2456.
35. Kaltenbach M, Tokuriki N. 2014. Dynamics and constraints of enzyme evolution. *J Exp Zool B Mol Dev Evol* 322:468-487.
36. Castillo Villamizar GA, Nacke H, Griese L, Taberner L, Funkner K, Daniel R. 2019. Characteristics of the first protein tyrosine phosphatase with phytase activity from a soil metagenome. *Genes* 10:101.
37. Bekalu ZE, Madsen CK, Dionisio G, Brinch-Pedersen H. 2017. *Aspergillus ficuum* phytase activity is inhibited by cereal grain components. *PLoS One* 12:e0176838.
38. Wyss M, Brugger R, Kronenberger A, Rémy R, Fimbel R, Oesterhelt G, Lehmann M, van Loon APGM. 1999. Biochemical characterization of fungal phytases (*myo*-inositol hexakisphosphate phosphohydrolases): catalytic properties. *Appl Environ Microbiol* 65:367-373.
39. Casey A, Walsh G. 2003. Purification and characterization of extracellular phytase from *Aspergillus niger* ATCC 9142. *Bioresour Technol* 86:183-188.

40. Monteiro PS, Guimarães VM, de Melo RR, de Rezende ST. 2015. Isolation of a thermostable acid phytase from *Aspergillus niger* UFV-1 with strong proteolysis resistance. *Braz J Microbiol* 46:251-260.
41. Singh B, Satyanarayana T. 2009. Characterization of a HAP-phytase from a thermophilic mould *Sporotrichum thermophile*. *Bioresour Technol* 100:2046-2051.
42. Alliegro MC. 2000. Effects of dithiothreitol on protein activity unrelated to thiol-disulfide exchange: for consideration in the analysis of protein function with Cleland's reagent. *Anal Biochem* 282:102-106.
43. Djoko KY, Achard MES, Phan M-D, Lo AW, Miraula M, Prombhul S, Hancock SJ, Peters KM, Sidjabat HE, Harris PN, Mitić N, Walsh TR, Anderson GJ, Shafer WM, Paterson DL, Schenk G, McEwan AG, Schembri MA. 2018. Copper ions and coordination complexes as novel carbapenem adjuvants. *Antimicrob Agents Chemother* 62:e02280-02217.
44. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol Ecol* 78:188-201.
45. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrupf M, Daniel R. 2016. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci Rep* 6:33696.
46. Villamizar GAC, Nacke H, Daniel R. 2017. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity. *Methods Mol Biol* 1539:249-260.
47. Kerovuo J, Lauraeus M, Nurminen P, Kalkkinen N, Apajalahti J. 1998. Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Appl Environ Microbiol* 64:2079-2085.
48. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257.
49. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA, Wagner L. 2003. Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31:28-33.
50. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944-945.
51. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y, Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H, Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA, Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC, Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto Silvio C E, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 45:D190-D199.
52. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmsberg W, Kapustin Y, Krasnov S, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Karsch-Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J. 2012. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 40:D13-D25.
53. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785-786.
54. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar GA, Pesseat S, Boland MA, Hunter FMI, Ten Hoopen P, Alako B, Amid C, Wilkinson DJ, Curtis TP, Cochrane G, Finn RD. 2018. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res* 46:D726-D735.
55. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

56. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870-1874.
57. Nei M, Saitou N. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425.
58. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242-W245.
59. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.
60. Anonymous. 2010. M9 minimal medium (standard). *Cold Spring Harb Protoc* 2010:pdb.rec12295.
61. Heinonen JK, Lahti RJ. 1981. A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase. *Anal Biochem* 113:313-317.
62. Gomori G. 1955. [16] Preparation of buffers for use in enzyme studies, p 138-146, *Methods Enzymol*, vol 1. Academic Press.

5.7. *Supplemental Information for Chapter 5*

Contents

Fig. S1. Protein sequence similarities of deduced gene products encoded by the insert of pLP05.

Fig. S2. Protein sequence similarities of deduced gene products encoded by the insert of pLP12.

Fig. S3. Krona chart showing the taxonomic affiliation of pLP05 and pLP12 inserts determined by Kaiju 1.6.3.

Fig. S4. Predicted structure of A) Mblp01 and B) Mblp02 by I-TASSER. The figure shows the best model of each enzyme (C-scores: -0.75 and -1.70, respectively)

Fig. S5. Effect of metal ions on the phytase activity of Mblp01 and Mblp02. The 100 % relative activities at concentrations of 1, 0.5 and 0.1 mM are 2.38, 2.61 and 2.77 for Mblp01, and 1.91, 2.03 and 1.94 U/mg for Mblp02, respectively.

Table S1: Effect of additives (1 mM) on Mblp01 and Mblp02 activity. The enzyme activities with phytate as substrate and without any added inhibitor of 2.31 (Mblp01) and 1.79 (Mblp02) U/mg were taken as 100% activity. Values are given as the mean of three experiments \pm standard deviations.

Supplementary Extended Data Table 1. UniProt codes, related proteins, genes and organisms of the sequences used for the construction of the phylogenetic tree depicted in Fig. 2.

ORF number	Position/strand	SmartBLAST result	ORF number	Position/strand	SmartBLAST result
<i>orf01</i>	1024 - 1125 (+)	PREDICTED: editase 1 isoform X3 (<i>Lepisosteus oculatus</i>)	<i>orf09</i>	2365 - 2487 (-)	No SmartBLAST hits found
<i>orf02</i>	1837 - 2532 (+)	No SmartBLAST hits found	<i>orf10</i>	2185 - 2343 (-)	protein trichome birefringence-like 14 (<i>Cucurbita moschata</i>)
<i>orf03</i>	680 - 1669 (+)	No SmartBLAST hits found	<i>orf11</i>	1789 - 1926 (-)	50S ribosomal protein L7/L12 (<i>Proteus mirabilis</i>)
<i>orf04</i>	1904 - 2080 (+)	No SmartBLAST hits found	<i>orf12</i>	1009 - 1089 (-)	winged helix-turn-helix domain-containing protein (<i>Bacillus cereus</i>)
<i>orf05</i>	2174 - 2530 (+)	LysR family (<i>Acidobacteria bacterium</i> 13_1_40CM_4_58_4)	<i>orf13</i>	598 - 750 (-)	hypothetical protein BO88DRAFT_401431 (<i>Aspergillus vadensis</i> CBS 113365)
<i>orf06</i>	135 - 233 (+)	DNA-binding protein (<i>Nostoc linckia</i>)	<i>orf14</i>	292 - 429 (-)	No SmartBLAST hits found
<i>orf07</i>	864 - 971 (+)	hypothetical protein SAMN05444955_1081 (<i>Lihuaxuella thermophila</i>)	<i>orf15</i>	1896 - 2078 (-)	No SmartBLAST hits found
<i>mblp01</i>	1158 - 2096 (+)	MBL fold metallo-hydrolase (<i>Verrucomicrobia bacterium</i>)	<i>orf16</i>	1182 - 1883 (-)	No SmartBLAST hits found
			<i>orf17</i>	276 - 392 (-)	hypothetical protein SCE1572_39240 (<i>Sorangium cellulosum</i> So0157-2)
			<i>orf18</i>	2153 - 2431 (-)	No SmartBLAST hits found
			<i>orf19</i>	1865 - 1957 (-)	No SmartBLAST hits found
			<i>orf20</i>	1355 - 1501 (-)	hypothetical protein MKAN_28760 (<i>Mycobacterium kansasii</i> ATCC 12478)
			<i>orf21</i>	392 - 1339 (-)	DUF72 domain-containing protein (<i>Acidobacteria bacterium</i>)
			<i>orf22</i>	2 - 157 (-)	response regulator (<i>Candidatus Koribacter versatilis</i>)

pLP05
(2496 bp)

Accession number
MH367836

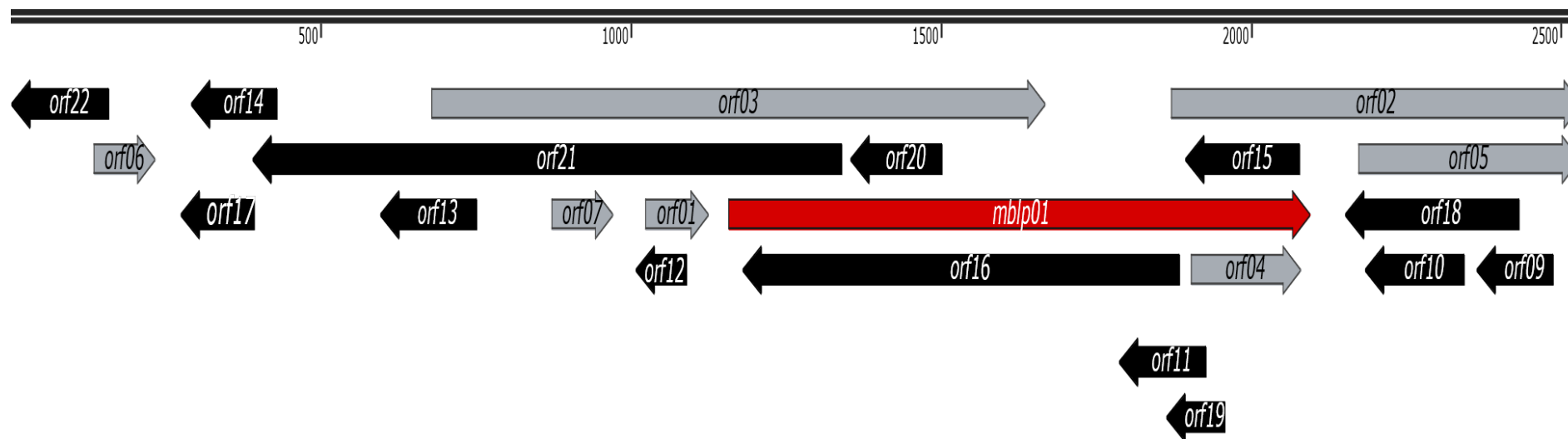


Fig. S1. Protein sequence similarities of deduced gene products encoded by the insert of pLP05.

pLP12
(5578 bp)
Accession number
MH367837

ORF number	Position/strand	SmartBLAST result	ORF number	Position/strand	SmartBLAST result
<i>orf01</i>	2050 - 2289 (+)	No SmartBLAST hits found	<i>orf21</i>	4754 - 4837 (-)	hypothetical protein (<i>Burkholderia</i> sp. CCGE1003)
<i>orf02</i>	2968 - 3069 (+)	hypothetical protein HSB1_40080 (<i>Halogranum salarium</i> B-1)	<i>orf22</i>	4418 - 4591 (-)	No SmartBLAST hits found
<i>orf03</i>	3103 - 3252 (+)	DUF1127 domain-containing protein (<i>Defluviimonas</i> sp. MS2-2)	<i>orf23</i>	4277 - 4381 (-)	DNA-binding protein (<i>Rhodothermus marinus</i>)
<i>orf04</i>	3727 - 3840 (+)	sulfatase (<i>Eisenbergiella massiliensis</i>)	<i>orf24</i>	2594 - 2845 (-)	unnamed protein product, partial (<i>Dibothriocephalus latus</i>)
<i>orf05</i>	4792 - 4899 (+)	peptidase M4 (<i>Burkholderiales</i> bacterium)	<i>orf25</i>	1625 - 2329 (-)	No SmartBLAST hits found
<i>orf06</i>	5179 - 5571 (+)	No SmartBLAST hits found	<i>orf26</i>	950 - 1048 (-)	hypothetical protein (<i>Nitrobacter vulgaris</i>)
<i>orf07</i>	197 - 463 (+)	No SmartBLAST hits found	<i>orf27</i>	839 - 949 (-)	No SmartBLAST hits found
<i>orf08</i>	1745 - 1888 (+)	MobA protein (<i>Methyloprofundus sedimenti</i>)	<i>orf28</i>	596 - 787 (-)	No SmartBLAST hits found
<i>orf09</i>	2129 - 2449 (+)	plasmid maintenance protein CcdB (<i>Sulfurirhabdus autotrophica</i>)	<i>orf29</i>	431 - 574 (-)	50S ribosomal protein L7ae, partial (<i>Candidatus Hydrothermae</i> bacterium)
<i>orf10</i>	2468 - 2716 (+)	No SmartBLAST hits found	<i>orf30</i>	2 - 151 (-)	LysR family transcriptional regulator (<i>Bacillus</i> sp. FIAT-27251)
<i>orf11</i>	3389 - 3565 (+)	DUF1013 domain-containing protein (<i>Phyllobacterium</i> sp. 1N-3)	<i>orf31</i>	4141 - 4263 (-)	No SmartBLAST hits found
<i>orf12</i>	99 - 398 (+)	Regulatory protein AfsR (<i>Paeniglutamicibacter gangotriensis</i> Lz1y)	<i>orf32</i>	2998 - 3144 (-)	hypothetical protein BCR33DRAFT_57825 (<i>Rhizoclostridium globosum</i>)
<i>mblp02</i>	510 - 1577 (+)	MBL fold metallo-hydrolase (<i>Sphingobium amiense</i>)	<i>orf33</i>	2266 - 2697 (-)	No SmartBLAST hits found
<i>orf14</i>	1662 - 2129 (+)	Protein CcdB (<i>Rhodocyclus bacterium</i> GWA2_65_19)	<i>orf34</i>	1954 - 2178 (-)	T9SS type A sorting domain-containing protein (<i>Taibaella soli</i>)
<i>orf15</i>	2739 - 2873 (+)	Antitoxin (Betaproteobacteria bacterium)	<i>orf35</i>	1612 - 1734 (-)	Polyketide synthase (<i>Sinorhizobium americanum</i>)
<i>orf16</i>	2892 - 3320 (+)	glycosyltransferase family 2 protein (<i>Sphingobium</i> sp. D43FB)	<i>orf36</i>	1144 - 1272 (-)	nitrate reductase gamma subunit (<i>Adlercreutzia</i> sp. Marseille-P7992)
<i>orf17</i>	3654 - 3749 (+)	cytochrome c (<i>Caballeronia zhejiangensis</i>)	<i>orf37</i>	622 - 723 (-)	No SmartBLAST hits found
<i>orf18</i>	3918 - 4385 (+)	No SmartBLAST hits found	<i>orf38</i>	2961 - 4205 (-)	No SmartBLAST hits found
<i>orf19</i>	4491 - 5147 (+)	Apurinic endonuclease-redox protein (<i>Symbiodinium microadriaticum</i>)	<i>orf39</i>	1707 - 1844 (-)	hypothetical protein AFLA70_98g003380 (<i>Aspergillus flavus</i> AF70)
<i>orf20</i>	5364 - 5627 (+)	No SmartBLAST hits found	<i>orf40</i>	1536 - 1628 (-)	MULTISPECIES: VIT family protein (<i>Chryseobacterium</i>)
			<i>orf41</i>	1218 - 1400 (-)	No SmartBLAST hits found
			<i>orf42</i>	741 - 884 (-)	PurS protein (<i>Sphingomonas aerea</i>)
			<i>orf43</i>	360 - 476 (-)	No SmartBLAST hits found

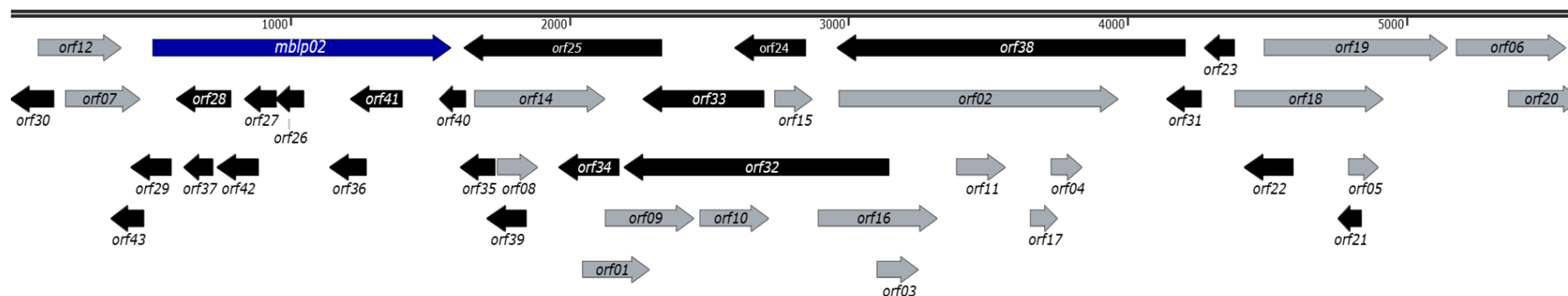


Fig. S2. Protein sequence similarities of deduced gene products encoded by the insert of pLP12.

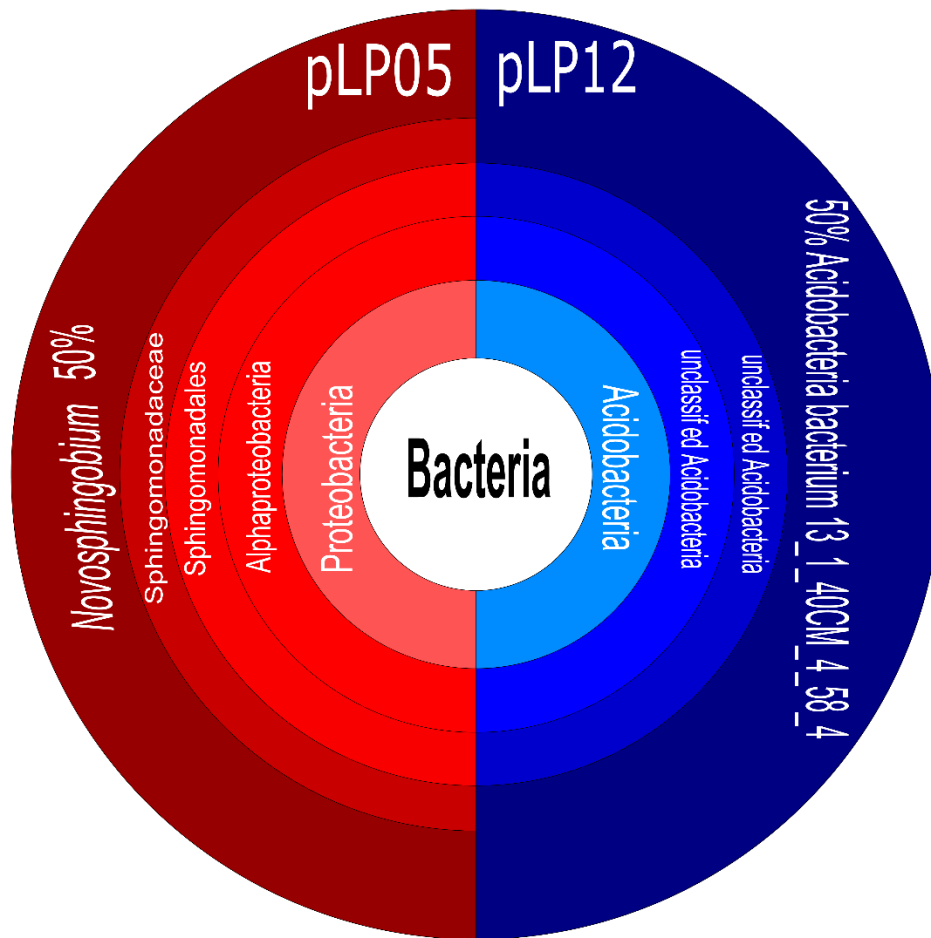


Fig. S3. Krona chart showing the taxonomic affiliation of pLP05 and pLP12 inserts determined by Kaiju 1.6.3.

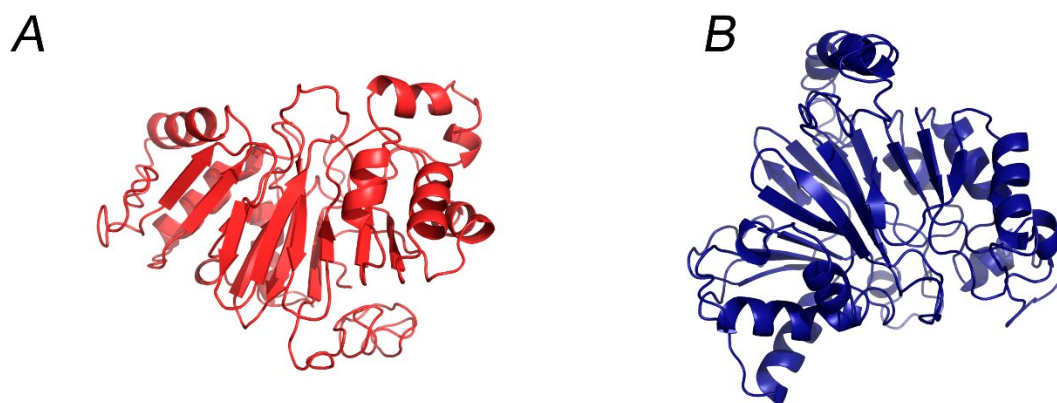


Fig. S4. Predicted structure of A) Mblp01 and B) Mblp02 by I-TASSER. The figure shows the best model of each enzyme (C-scores: -0.75 and -1.70, respectively)

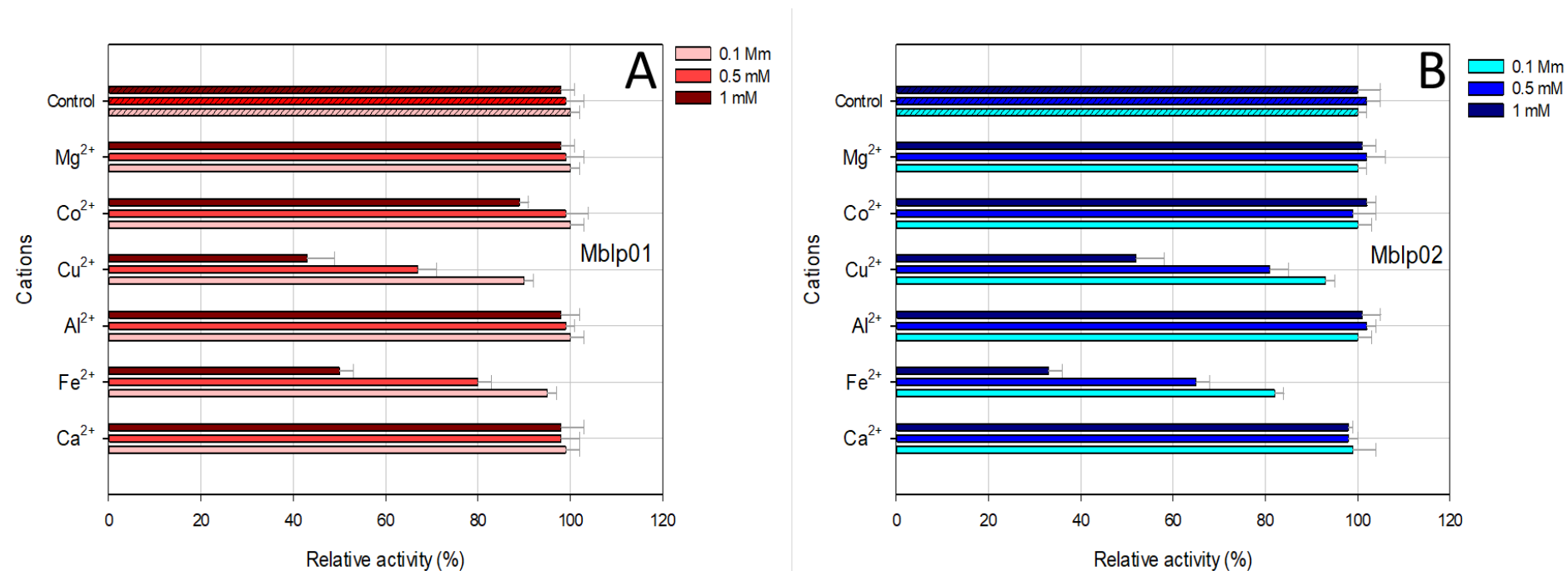


Fig. S5. Effect of metal ions on the phytase activity of Mblp01 and Mblp02. The 100 % relative activities at concentrations of 1, 0.5 and 0.1 mM are 2.38, 2.61 and 2.77 for Mblp01, and 1.91, 2.03 and 1.94 U/mg for Mblp02, respectively.

Table S1: Effect of additives (1 mM) on Mblp01 and Mblp02 activity. The enzyme activities with phytate as substrate and without any added inhibitor of 2.31 (Mblp01) and 1.79 (Mblp02) U/mg were taken as 100% activity. Values are given as the mean of three experiments \pm standard deviations.

Inhibitors	Phytase relative activity (%)	
	Mblp01	Mblp02
Control*	100 \pm 0.83	100 \pm 0.69
EDTA	50.38 \pm 1.12	62.85 \pm 1.46
Tungstate	45.52 \pm 2.53	22.05 \pm 2.33
Oxalate	69.19 \pm 1.98	35.56 \pm 3.02
SDS	N.D	2.92 \pm 1.61
DTT	7.21 \pm 0.92	14.02 \pm 2.51

*The enzyme activity toward phytate without any added inhibitor was taken as 100% activity, 2.31 and 1.79 U/mg for Mblp01 and Mblp02, respectively. Values are given as the mean of three experiments \pm standard deviations. N.D: Not detected.

Supplementary Extended Data Table 1. UniProt codes, related proteins, genes and organisms of the sequences used for the construction of the phylogentic tree depicted in Fig. 2.

Fasta-Name	Protein	Gene	Organism
A1WQ33	Coenzyme PQQ synthesis protein B	pqqB	<i>Verminephrobacter eiseniae</i> (strain EF01-2)
A2SJ00	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylibium petroleiphilum</i> (strain ATCC BAA-1232 / LMG 22953 / PM1)
A3DK71	Beta-lactamase superfamily hydrolase	Cthe_3155	<i>Clostridium thermocellum</i> (strain ATCC 27405 / DSM 1237 / NBRC 103400 / NCIMB 10682 / NRRL B-4536 / VPI 7372) (<i>Ruminiclostridium thermocellum</i>)
A4AUL5	Putative hydrolase	FB2170_12641	<i>Maribacter</i> sp. (strain HTCC2170 / KCCM 42371)
A0JVX8	Zn-dependent hydrolases of the beta-lactamase fold-like protein	Arth_1808	<i>Arthrobacter</i> sp. (strain FB24)
A4X5M1	Zn-dependent hydrolase of the beta-lactamase fold-like protein	Strop_1707	<i>Salinispora tropica</i> (strain ATCC BAA-916 / DSM 44818 / CNB-440)
A4YZY4	Coenzyme PQQ synthesis protein B	pqqB	<i>Bradyrhizobium</i> sp. (strain ORS 278)
A7H951	Uncharacterized protein	Anae109_1038	<i>Anaeromyxobacter</i> sp. (strain Fw109-5)
A9GAR8	Uncharacterized protein	sce2742	<i>Sorangium cellulosum</i> (strain So ce56) (<i>Polyangium cellulosum</i> (strain So ce56))
A9GHH6	Uncharacterized protein	sce9325	<i>Sorangium cellulosum</i> (strain So ce56) (<i>Polyangium cellulosum</i> (strain So ce56))
B0UGZ8	Beta-lactamase domain protein	M446_2411	<i>Methylobacterium</i> sp. (strain 4-46)
B0VI78	UPF0173 metal-dependent hydrolase CLOAM1192	CLOAM1192	<i>Cloacimonas acidaminovorans</i> (strain Evry)
B1LUZ1	Beta-lactamase domain protein	Mrad2831_3601	<i>Methylobacterium radiotolerans</i> (strain ATCC 27329 / DSM 1819 / JCM 2831 / NBRC 15690 / NCIMB 10815 / 0-1)
B2IDU9	Coenzyme PQQ synthesis protein B	pqqB	<i>Beijerinckia indica</i> subsp. <i>indica</i> (strain ATCC 9039 / DSM 1715 / NCIB 8712)
B3DX71	Zn-dependent hydrolase of the beta-lactamase fold	Minf_1726	<i>Methylacidiphilum infernorum</i> (isolate V4) (<i>Methylokorus infernorum</i> (strain V4))
B6IP37	Uncharacterized protein	RC1_2067	<i>Rhodospirillum centenum</i> (strain ATCC 51521 / SW)
B6YQL0	Uncharacterized protein	CFPG_219	<i>Azobacteroides pseudotriconymphae</i> genomovar. CFP2
B9JKU6	Uncharacterized protein	Arad_9513	<i>Agrobacterium radiobacter</i> (strain K84 / ATCC BAA-868)
C1F564	UPF0173 metal-dependent hydrolase ACP_3152	ACP_3152	<i>Acidobacterium capsulatum</i> (strain ATCC 51196 / DSM 11244 / JCM 7670 / NBRC 15755 / NCIMB 13165 / 161)
C1F6C1	Metallo-beta-lactamase family protein	ACP_3353	<i>Acidobacterium capsulatum</i> (strain ATCC 51196 / DSM 11244 / JCM 7670 / NBRC 15755 / NCIMB 13165 / 161)
C6WVF4	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylotenera mobilis</i> (strain JLW8 / ATCC BAA-1282 / DSM 17540)
C6XDZ8	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylovorus glucosetrophus</i> (strain SIP3-4)
C7JE57	Coenzyme PQQ synthesis protein B	pqqB	<i>Acetobacter pasteurianus</i> (strain NBRC 3283 / LMG 1513 / CCTM 1153)
D0LX98	Beta-lactamase domain protein	Hoch_3638	<i>Haliangium ochraceum</i> (strain DSM 14365 / JCM 11303 / SMP-2)
D2AUV7	Uncharacterized protein	Sros_3902	<i>Streptosporangium roseum</i> (strain ATCC 12428 / DSM 43021 / JCM 3005 / NI 9100)

D2S3Z3	Beta-lactamase domain protein	Gobs_0215	<i>Geodermatophilus obscurus</i> (strain ATCC 25078 / DSM 43160 / JCM 3152 / G-20)
D5QAQ5	Coenzyme PQQ synthesis protein B	pqqB	<i>Komagataeibacter hansenii</i> ATCC 23769
D5V712	Coenzyme PQQ synthesis protein B	pqqB	<i>Arcobacter nitrofigilis</i> (strain ATCC 33309 / DSM 7299 / LMG 7604 / NCTC 12251 / CI) (<i>Campylobacter nitrofigilis</i>)
D5XBH6	Metal-dependent hydrolase	TherJR_2568	<i>Thermincola potens</i> (strain JR)
D6YB30	Beta-lactamase domain protein	Tbis_1677	<i>Thermobispora bispora</i> (strain ATCC 19993 / DSM 43833 / CBS 139.67 / JCM 10125 / NBRC 14880 / R51)
D6ZD67	Metallo-beta-lactamase family protein	Srot_2823	<i>Segniliparus rotundus</i> (strain ATCC BAA-972 / CDC 1076 / CIP 108378 / DSM 44985 / JCM 13578)
D7BYX6	Uncharacterized protein	SBI_02564	<i>Streptomyces bingchenggensis</i> (strain BCW-1)
D7DHP1	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylotenera versatilis</i> (strain 301)
E0SIG9	Outer membrane protein RomA	Dda3937_02835	<i>Dickeya dadantii</i> (strain 3937) (<i>Erwinia chrysanthemi</i> (strain 3937))
E2PX85	Zn-dependent hydrolase	SCLAV_5089	<i>Streptomyces clavuligerus</i> (strain ATCC 27064 / DSM 738 / JCM 4710 / NBRC 13307 / NCIMB 12785 / NRRL 3585 / VKM Ac-602)
E3FNR1	Conserved uncharacterized protein	STAUR_0331	<i>Stigmatella aurantiaca</i> (strain DW4/3-1)
E4RTV2	Beta-lactamase domain protein	Lbys_2998	<i>Leadbetterella byssophila</i> (strain DSM 17132 / KACC 11308 / 4M15)
E6SEG1	Uncharacterized protein	Intca_3356	<i>Intrasporangium calvum</i> (strain ATCC 23552 / DSM 43043 / JCM 3097 / NBRC 12989 / 7 KIP)
E8V4I3	Uncharacterized protein	AcipR4_4058	<i>Terriglobus saanensis</i> (strain ATCC BAA-1853 / DSM 23119 / SP1PR4)
E9T554	Metallo-beta-lactamase domain protein	HMPREF0724_13682	<i>Rhodococcus hoagii</i> ATCC 33707
F0S2Z7	Metallo-beta-lactamase family protein	Dester_0568	<i>Desulfurobacterium thermolithotrophum</i> (strain DSM 11699 / BSA)
F2I749	Metallo-beta-lactamase domain protein	HMPREF9243_1704	<i>Aerococcus urinae</i> (strain ACS-120-V-Col10a)
F2KPP9	UPF0173 metal-dependent hydrolase Arcve_1577	Arcve_1577	<i>Archaeoglobus veneficus</i> (strain DSM 11195 / SNP6)
F4B347	Beta-lactamase domain-containing protein	Krodi_2683	<i>Dokdonia</i> sp. (strain 4H-3-7-5) (<i>Krokinobacter</i> sp. (strain 4H-3-7-5))
F4CM85	Coenzyme PQQ synthesis protein B	pqqB	<i>Pseudonocardia dioxanivorans</i> (strain ATCC 55486 / DSM 44775 / JCM 13855 / CB1190)
F4CPX3	Beta-lactamase domain protein	Psed_3990	<i>Pseudonocardia dioxanivorans</i> (strain ATCC 55486 / DSM 44775 / JCM 13855 / CB1190)
F4F180	Zn-dependent hydrolase of the beta-lactamase fold-like protein	VAB18032_14020	<i>Verrucosispora maris</i> (strain AB-18-032)
F4GHG2	Beta-lactamase domain protein	Spico_1345	<i>Sphaerochaeta coccoides</i> (strain ATCC BAA-1237 / DSM 17374 / SPN1) (<i>Spirochaeta coccoides</i>)
F4PAY0	Uncharacterized protein	BATDEDRAFT_13909	<i>Batrachochytrium dendrobatidis</i> (strain JAM81 / FGSC 10211) (<i>Frog chytrid fungus</i>)
F6EMB1	Uncharacterized protein	AS9A_0865	<i>Hoyosella subflava</i> (strain DSM 45089 / JCM 17490 / NBRC 109087 / DQS3-9A1) (<i>Amycolicoccus subflavus</i>)
F7QLQ9	Coenzyme PQQ synthesis protein B	pqqB	<i>Bradyrhizobiaceae bacterium SG-6C</i>

F8JZ7	Uncharacterized protein	SCATT_49660	<i>Streptomyces cattleya</i> (strain ATCC 35852 / DSM 46488 / JCM 4925 / NBRC 14057 / NRRL 8057)
G0L7Q9	Metallo-beta-lactamase superfamily protein	zobellia_4075	<i>Zobellia galactanivorans</i> (strain DSM 12802 / CCUG 47099 / CIP 106680 / NCIMB 13871 / Dsij)
G2ITI4	Putative hydrolase	SLG_20190	<i>Sphingobium</i> sp. (strain NBRC 103272 / SYK-6)
G2IZA3	Coenzyme PQQ synthesis protein B	pqqB	<i>Pseudogulbenkiania</i> sp. (strain NH8B)
G2PKK9	Beta-lactamase domain protein	Murru_3034	<i>Muricauda ruestringensis</i> (strain DSM 13258 / CIP 107369 / LMG 19739 / B1)
G4R899	Coenzyme PQQ synthesis protein B	pqqB	<i>Pelagibacterium halotolerans</i> (strain DSM 22347 / JCM 15775 / CGMCC 1.7692 / B2)
G4RKZ5	Ribonuclease Z	rnz	<i>Thermoproteus tenax</i> (strain ATCC 35583 / DSM 2078 / JCM 9277 / NBRC 100435 / Kra 1)
G7DVV2	Uncharacterized protein	Mo01365	<i>Mixia osmundae</i> (strain CBS 9802 / IAM 14324 / JCM 22182 / KY 12970)
G7ZEV0	Coenzyme PQQ synthesis protein B	pqqB	<i>Azospirillum lipoferum</i> (strain 4B)
G8NUJ1	Beta-lactamase domain protein	AciX8_2113	<i>Granulicella mallensis</i> (strain ATCC BAA-1857 / DSM 23137 / MP5ACTX8)
G8NYQ4	Beta-lactamase	AciX8_0109	<i>Granulicella mallensis</i> (strain ATCC BAA-1857 / DSM 23137 / MP5ACTX8)
G8POA8	Coenzyme PQQ synthesis protein B	pqqB	<i>Granulicella mallensis</i> (strain ATCC BAA-1857 / DSM 23137 / MP5ACTX8)
G8PVE1	Zn-dependent hydrolase of the beta-lactamase fold-like protein	PSE_p0147	<i>Pseudovibrio</i> sp. (strain FO-BEG1)
H0Q3B3	Coenzyme PQQ synthesis protein B	pqqB	<i>Azoarcus</i> sp. KH32C
H2JS06	Uncharacterized protein	SHJG_7667	<i>Streptomyces hygroscopicus</i> subsp. <i>jinggangensis</i> (strain 5008)
H6QY58	Putative Zn-dependent hydrolase of the beta-lactamase fold	NOCYR_0692	<i>Nocardia cyriacigeorgica</i> (strain GUH-2)
H8MFU1	Uncharacterized protein	COCOR_07021	<i>Coralloccoccus coralloides</i> (strain ATCC 25202 / DSM 2259 / NBRC 100086 / M2) (<i>Myxococcus coralloides</i>)
I2EZN2	Uncharacterized protein	Emtol_4010	<i>Emticia oligotrophica</i> (strain DSM 17448 / GPTSA100-15)
I3TIV4	Beta-lactamase superfamily hydrolase	TMO_0853	<i>Tistrella mobilis</i> (strain KA081020-065)
I3TL82	Metallo-beta-lactamase superfamily	lipB	<i>Tistrella mobilis</i> (strain KA081020-065)
I3ZD26	Putative Zn-dependent hydrolase of beta-lactamase fold protein	Terro_0812	<i>Terriglobus roseus</i> (strain DSM 18391 / NRRL B-41598 / KBS 63)
I4D0N2	Putative Zn-dependent hydrolase of beta-lactamase fold protein	Desaci_0284	<i>Desulfosporosinus acidiphilus</i> (strain DSM 22704 / JCM 16185 / SJ4)
K7SBP4	Coenzyme PQQ synthesis protein B	pqqB	<i>Gluconobacter oxydans</i> H24
M1FI32	Coenzyme PQQ synthesis protein B	pqqB	<i>Marinobacter</i> sp. BSs20148
PQQB_AZOVD	Coenzyme PQQ synthesis protein B	pqqB	<i>Azotobacter vinelandii</i> (strain DJ / ATCC BAA-1303)
PQQB_GLU DA	Coenzyme PQQ synthesis protein B	pqqB	<i>Gluconacetobacter diazotrophicus</i> (strain ATCC 49037 / DSM 5601 / PAI5)
PQQB_GLUOX	Coenzyme PQQ synthesis protein B	pqqB	<i>Gluconobacter oxydans</i> (strain 621H) (<i>Gluconobacter suboxydans</i>)

PQQB_METCA	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylococcus capsulatus</i> (strain ATCC 33009 / NCIMB 11132 / Bath)
PQQB_METEA	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylobacterium extorquens</i> (strain ATCC 14718 / DSM 1338 / JCM 2805 / NCIMB 9133 / AM1) (<i>Methylorubrum extorquens</i>)
PQQB_METFK	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylobacillus flagellatus</i> (strain KT / ATCC 51484 / DSM 6875)
PQQB_METS4	Coenzyme PQQ synthesis protein B	pqqB	<i>Methylobacterium</i> sp. (strain 4-46)
PQQB_PSEAE	Coenzyme PQQ synthesis protein B	pqqB	<i>Pseudomonas aeruginosa</i> (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1)
PQQB_PSEPK	Coenzyme PQQ synthesis protein B	pqqB	<i>Pseudomonas putida</i> (strain ATCC 47054 / DSM 6125 / NCIMB 11950 / KT2440)
PQQB_RHOPA	Coenzyme PQQ synthesis protein B	pqqB	<i>Rhodopseudomonas palustris</i> (strain ATCC BAA-98 / CGA009)
PQQB_XANCP	Coenzyme PQQ synthesis protein B	pqqB	<i>Xanthomonas campestris</i> pv. <i>campestris</i> (strain ATCC 33913 / DSM 3586 / NCPPB 528 / LMG 568 / P 25)
Q027I3	Beta-lactamase domain protein	Acid_1840	<i>Solibacter usitatus</i> (strain Ellin6076)
Q0BT75	Metal-dependent hydrolase	GbCGDNIH1_1079	<i>Granulibacter bethesdensis</i> (strain ATCC BAA-1260 / CGDNIH1)
Q21AY6	Coenzyme PQQ synthesis protein B	pqqB	<i>Rhodopseudomonas palustris</i> (strain BisB18)
Q2K048	Hypothetical conserved protein	RHE_PE00054	<i>Rhizobium etli</i> (strain CFN 42 / ATCC 51251)
Q2RTP6	Beta-lactamase-like	Rru_A1699	<i>Rhodospirillum rubrum</i> (strain ATCC 11170 / ATH 1.1.1 / DSM 467 / LMG 4362 / NCIB 8255 / S1)
Q2W4H5	Metal-dependent hydrolase of the beta-lactamase superfamily I	amb2446	<i>Magnetospirillum magneticum</i> (strain AMB-1 / ATCC 700264)
Q3A832	Metal-dependent hydrolase, beta-lactamase superfamily	Pcar_0199	<i>Pelobacter carbinolicus</i> (strain DSM 2380 / NBRC 103641 / GraBd1)
Q5JJ51	Predicted hydrolase, metallo-beta-lactamase superfamily	TK1778	<i>Thermococcus kodakarensis</i> (strain ATCC BAA-918 / JCM 12380 / KOD1) (<i>Pyrococcus kodakaraensis</i> (strain KOD1))
Q6LL65	Uncharacterized protein	RSP1122	<i>Photobacterium profundum</i> (strain SS9)
Q82LZ2	Uncharacterized protein	SAVERM_1868	<i>Streptomyces avermitilis</i> (strain ATCC 31267 / DSM 46492 / JCM 5070 / NBRC 14893 / NCIMB 12804 / NRRL 8165 / MA-4680)
R4LT43	Zn-dependent hydrolase	L083_6641	<i>Actinoplanes</i> sp. N902-109
R5AH50	Beta-lactamase-like protein	BN454_01422	<i>Clostridium</i> sp. CAG:1024
R5CAE1	Metallo-beta-lactamase domain protein	BN727_01472	<i>Bacteroides</i> sp. CAG:598
R5CC61	Metallo-beta-lactamase domain protein	BN458_01277	<i>Prevotella</i> sp. CAG:1058
R5KJV9	Metallo-beta-lactamase domain protein	BN467_02063	<i>Prevotella</i> sp. CAG:1124
R5NLD0	Metallo-beta-lactamase domain protein	BN471_01507	<i>Paraprevotella clara</i> CAG:116
R6PKU0	Lipoate-protein ligase B	BN637_01418	<i>Prevotella</i> sp. CAG:386
R6VV32	Lipoate-protein ligase B	BN673_02141	<i>Prevotella</i> sp. CAG:474
R6W9B0	Putative phnP protein	BN725_00349	<i>Prevotella</i> sp. CAG:592

R6ZJ2	Uncharacterized protein	BN815_01976	<i>Firmicutes bacterium CAG:94</i>
U6B578	PhnP protein	phnP	<i>Candidatus Liberibacter americanus str. Sao Paulo</i>
V6KSG9	Membrane protein	M878_07850	<i>Streptomyces roseochromogenus subsp. oscitans DS 12.976</i>
W0AF05	Uncharacterized protein	NX02_12760	<i>Sphingomonas sanxanigenens DSM 19645 = NX02</i>
WORD40	Uncharacterized protein	J421_0825	<i>Gemmatirosa kalamazoonesis</i>
W5WH03	Uncharacterized protein	KALB_4085	<i>Kutzneria albida DSM 43870</i>
W8F114	Uncharacterized protein	Hsw_2292	<i>Hymenobacter swuensis DY53</i>
Y1100_PSELT	UPF0173 metal-dependent hydrolase Tlet_1100	Tlet_1100	<i>Pseudothermotoga lettingae (strain ATCC BAA-301 / DSM 14385 / NBRC 107922 / TMO) (Thermotoga lettingae)</i>
Y1493_THEPD	UPF0173 metal-dependent hydrolase Tpen_1493	Tpen_1493	<i>Thermofilum pendens (strain DSM 2475 / Hrk 5)</i>
Y2542_CHLAA	UPF0173 metal-dependent hydrolase Caur_2542	Caur_2542	<i>Chloroflexus aurantiacus (strain ATCC 29366 / DSM 635 / J-10-fl)</i>
Y3308_DESVH	UPF0173 metal-dependent hydrolase DVU_3308	DVU_3308	<i>Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / DSM 644 / NCIMB 8303)</i>
P0A8V0	Ribonuclease BN	rbn	<i>Escherichia coli (strain K12)</i>
P0AC84	Hydroxyacylglutathione hydrolase GloB	gloB	<i>Escherichia coli (strain K12)</i>
P16692	Phosphoribosyl 1,2-cyclic phosphate phosphodiesterase	phnP	<i>Escherichia coli (strain K12)</i>
P22434	3',5'-cyclic-nucleotide phosphodiesterase 1	PDE1	<i>Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)</i>
Q05584	Hydroxyacylglutathione hydrolase, cytoplasmic isozyme	GLO2	<i>Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)</i>
Q12320	Hydroxyacylglutathione hydrolase, mitochondrial	GLO4	<i>charomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)</i>
S5MMG8	Alkaline phytase	phy	<i>Bacillus subtilis</i>
A4UU76	3-phytase	phyL	<i>Bacillus licheniformis</i>
AAQ13669.1	Myo-inositol hexaphosphate phosphohydrolase	phyA	<i>Selenomonas ruminantium</i>
ABC69367.2	Protein tyrosine phosphatase-like inositol polyphosphate phosphatase	phyAsI	<i>Selenomonas lactificex</i>
AHA61669.1	Histidine acid phosphatase phyt	phyA	<i>Thielavia heterothallica (Myceliophthora thermophila)</i>
ABI95370.1	Phytase		<i>Yersinia intermedia</i>
ABP96799.1	Purple acid phosphatase	PAP	<i>Nicotiana tabacum (Common tobacco)</i>

6 General Discussion

6.1. *Soil as Source of Metagenome-Derived Biocatalysts*

Soils are heterogeneous environments that contain large numbers of different ecological niches and correspondingly diverse soil microbial communities. The structure and diversity of these communities vary with soil type, soil depth, intrinsic physicochemical soil characteristics, climate and biotic factors such as soil-associated tree species (1) (Appendix 8.2.1). Soil represents an important source of biomolecules with significant biotechnological and scientific value. As ecosystem soil hosts a large number of complex processes and interactions among eukaryotes and prokaryotes (1, 2). Plenty of those processes and interactions are mediated by enzymes participating in diverse pathways, which in many cases are restricted to the vast majority of so far uncultured microorganisms. In this study, soil samples from the German Biodiversity Exploratories Schorfheide-Chorin (northeastern Germany), Hainich-Dün (central Germany), and Schwäbische Alb (southwestern Germany) (3) were used as starting material for the development of a new function-based metagenomic method for the retrieval of new phosphatases. The soil matrixes of the German Biodiversity Exploratories represented the perfect starting point for our studies as the microbial communities associated to the soil habitats has been well studied and based on functional predictions derived from community composition the presence of putative genes encoding alkaline and acid phosphatases associated to phosphorus cycling was indicated (1, 4, 5).

Metagenomics contributed to identification of new functional enzymes related to important cycles and pathways. Mining of soil metagenomes has enabled the exploration and exploitation of different enzyme types. The global metagenomic search for enzymes has focused on enzymes mainly with application potential, e.g. esterases, lipases, glycosyl hydrolases and lactonases followed by oxidoreductases, including monooxygenases and dioxygenases (6). Recent reviews and meta-analysis consider few enzyme types as preferred targets of metagenomic surveys with industrial interest i.e. acylases, nitrilases, phosphatases, proteases, oxidoreductases, glycosyl hydrolases and lipases/esterases (7, 8). Nevertheless, the effectivity and frequency of retrieving active clones and subsequent enzymatic characterization differs largely among the different targets, depending among other things of the abundance of the enzymatic activities across the organisms (9).

6.2. *Phosphatases in Metagenome Surveys*

With the rise of next generation sequencing methods, thousands of microorganisms from very diverse environments have been sequenced and their genomes analyzed (10, 11) (Appendix 8.2.2, 8.2.3). Similarity searches revealed the presence of genes homologous to known phosphatases in the genomes and metagenomes from very diverse habitats (9, 12, 13).

Despite, the frequent metagenomic surveys and their key roles in many cellular and biotechnological processes the amount of functionally characterized phosphatases derived from metagenome approaches is still low (8). Ferrer et al. 2016 analyzed the last 20 years of enzyme-related metagenomic data represented by 2,192 different sites distributed across the planet (7). Extensive multi-environment meta-analysis covered 6,100 clones exhibiting enzymes activities, which were isolated and characterized (fully or partially). In the specific case of phosphatases that survey did not show a large proportion of characterized enzymes. (Fig. 1).

This study adds 23 new clones and 33 new genes associated to phosphatase activity. Phosphatases can be classified by their structural folds and in some cases for their biochemical properties and biological functions (14). Among the gene products encoded by the 33 candidate genes, alkaline phosphatases were identified as the most abundant group with five representatives. Histidine phosphatases and phospholipases were represented with four proteins each. Phosphoserine-phosphatases and protein-tyrosine phosphatases were represented by three genes each while the acid phosphatases and metallo- β -lactamases with phosphatase activity were encoded by two genes in each case. Finally, one encoded protein was associated to a mismatch repair ATPase.

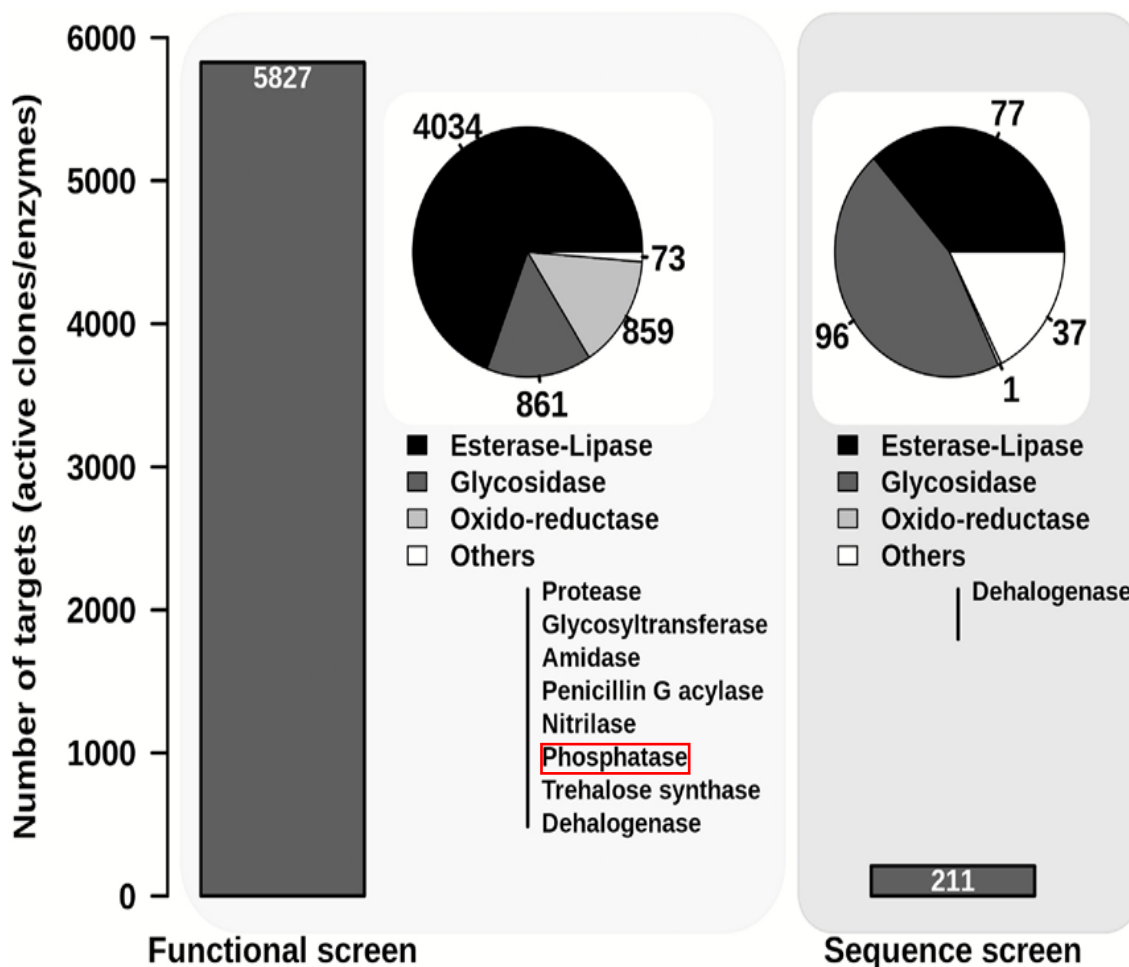


Fig. 1. Survey of the total number of targets (clones and/or single enzymes and/or sequences encoding enzymes) identified by metagenomic studies. The distribution of selected targets by type of enzyme activity is shown for both screening methods used: function-based and sequence-based screening. The figure is based on studies that were published over the last two decades using native (left) and sequence-based (right) screening protocols. The databases used to provide such estimations were SCOPUS, PubMed, WOK and the IMG/M of the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) and UniProtKB/Swiss-Prot. Modified from (7).

6.3. Metagenomics for the Recovery of New Phytases

Within the phosphatases, phytases are of special interest because in addition to their relatively poorly studied *in vivo* functions, commercialized phytases have a growing international market and are one of the most important type of biocatalyst for the global enzyme industry (15). With very few exceptions, which includes the five enzymes described in the chapters 3, 4 and 5 of this study most characterized phytate degrading enzymes are derived exclusively from cultured individual microorganisms. Consequently, the real diversity of phytate-degrading enzymes remains underestimated. This partially limits our understanding of the biological functions of phytate degrading enzymes and also keeps concealed interesting enzymes with putative biotechnological relevance. In

contrast to other biocatalyst such as lipases or cellulases, metagenome-derived phytases have not been commercialized so far (16).

Several groups around the world have made efforts aiming to retrieve phytase homologous genes from the unculturable fraction of microorganisms. The initial attempts in this direction were performed based on sequence similarity, mostly by using PCR amplification with degenerated primers. The designed primers were based on sequences of previously reported phytases from culturable microorganisms. This method has two major disadvantages. First, it limits largely the recovered genes to close relatives of the preceding genes (17-19). Moreover, it is uncertain if the identified and subsequently cloned gene will encode a functional enzyme capable of using phytate as substrate. This is due to the fact that almost all types of phytases are part of larger phosphatases superfamilies, which are capable of processing phosphorylated substrates but not necessarily phytate (20).

The subsequent improvement was the combination of PCR-based methods with large-scale sequencing experiments. The sequence information is used to design primers and other probes which are suitable to recover full-length versions of specific target genes (21). This type of procedure is called gene-targeted-metagenomics and was first used for the recovery of genes encoding aromatic dioxygenases from contaminated soil samples. Originally, gene-targeted-metagenomics utilized pyrosequencing and required that the targeted gene contained enough conserved regions of suitable distance for PCR. Without those conditions, the designed sets of primers did not had the essential coverage for the amplification of the genes of interest (22). Nowadays, by combining shotgun sequencing methods and bioinformatics, these limitations are ovoid. Under the existent methodologies, all DNA is extracted and subsequently sheared into smaller fragments for library preparation which allows independent sequencing of the fragments. The resulting DNA sequences (i.e., reads) can be mined for genes with specific biological functions (23).

In the specific case of phytases, few sequence-based metagenomic surveys have been employed. Most of them focused on mining metagenome-derived gene sequences from sources such as subsurface groundwater, acidic peat-soil microbiomes, rumen and insect-cultivated fungus-gardens. The resulting phytases were characterized and exhibited habitat-related characteristics such as thermostability and acid resilience(24-26). The major limitation of those approaches is the low performance identifying novel phytase-encoding genes with no or low similarity to known ones. Mootapally et al. (26) identified

a HAPhy by mining the metagenomic data obtained from the rumen of buffalos. However, the identified phytase showed more than 95% sequence identity to a previously reported phytase encoded by species belonging to genus *Prevotella* (26). Although, other sequence-based attempts reported lower sequence identity with known phytases ($\approx 50\%$) the recovered phytases were again strictly limited to a typical type of known phytases i.e. HAPhys (25). Thus, the scavenge of phytases by sequence-driven approaches can provide results but remains limited to recover only enzymes harboring known molecular signatures.

Consequently, function-based approaches are preferred over the sequence-based methods to analyze the diversity of protein families with specific functions and for the discovering genes with novel functions (21). In the case of the phytases, function-driven screenings are also required to encounter “real” novel genes encoding phytases.

Different function driven approaches are broadly used to recover novel biocatalysts e.g. heterologous complementation, induced gene expression and phenotypical detection of the activities (21, 27). Heterologous complementation of host strains or mutants is a simple fast and highly selective method in which host strains or their mutants require the expression of the targeted genes for growth under selective conditions. Different types of genes and biocatalysts have been recovered by this method (e.g. DNA polymerases and RNAses) (28, 29). Conversely, induced gene expression does not limit the growth of the clones carrying the library, but instead this method involves the use of different strategies for the detection of genes of interest. The strategies are based on the coupled expression of those genes with reporter genes e.g. *gfp*. Compared with the screening by heterologous complementation, the induced gene expression might lead to a higher rate of false positives, because the transcriptional activation of the reporter genes could be occasionally induced by cellular effectors and not necessarily by specific substrates (21). Finally, one of the most broadly used function-based screenings is the phenotypical detection of the target activities. This method utilizes chemical dyes or chromophores derivated of enzyme substrates for the activity detection (27). These substances are incorporated into the growth medium, allowing the recognition of individual clones bearing the specific targeted metabolic capabilities (21). Plenty of successful examples of the application of this approach are reported in the literature e.g. proteases, esterases, lipases, cellulases among many others (30-33).

The register of the positive clones/activities derived from function-based screening strategies is achieved by different methods. Those methods have variable levels of technological requirements from common Petri dishes to coupled droplet-laser devices. Fig. 2 shows a simplified overview of the library construction steps and different methods for the detection of the target activities in function-based metagenomics (27).

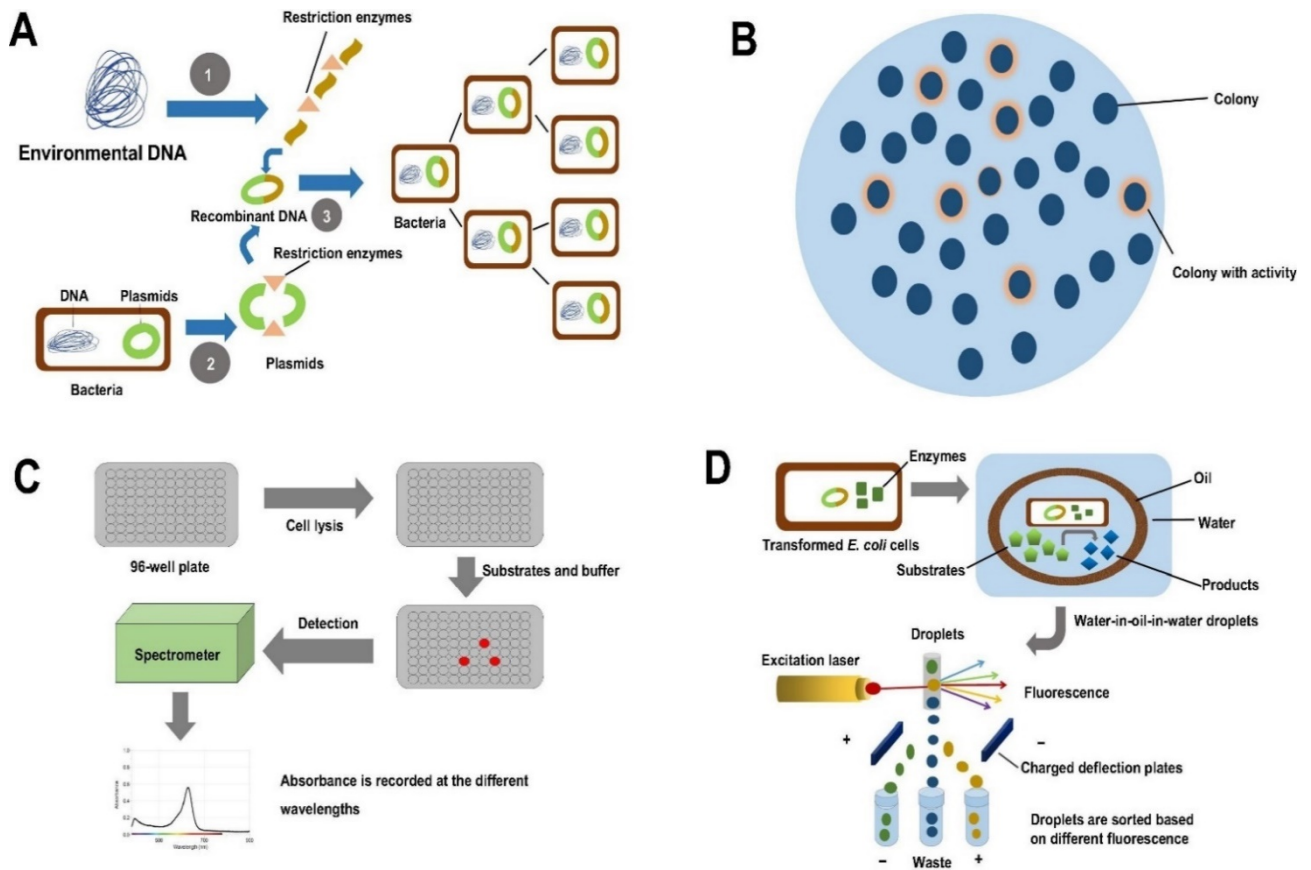


Fig. 2. Simplified overview of function-based metagenomic screening procedures. **A.** Steps involved in the construction of a small DNA fragment library from environmental metagenome or from culturable bacteria. **B.** Agar plate activity screening. **C.** Microtiter plate screening. **D.** Microfluidics coupled with fluorescence-activated cell sorting. Modified from (27).

Different from other relevant biocatalysts such as lipases or glycosidases, phytases derived from function-based screenings are barely reported. Although, some progress has been made in order to get phytases from metagenomes, a large gap of information still exists. One of the main hindrances is the lack of reliable screening strategies. To my knowledge besides this study only two other published reports have pursued the obtention of phytate-degrading enzymes by using function-based approaches (34, 35).

6.4. A Simple and Effective Function-Based Method for the Retrieval of Novel Phosphatases/Phytases

One of the first function-based strategies for the search of phytases was performed by Tan et al. 2014 (34). A metagenomic library cloned in a fosmid vector, (pCC1FOS) was function-driven screened. The metagenomic library was derived from cultivated farm soil samples. That study identified initially 28 candidate clones capable to grow on a selective medium, containing phytate as the sole carbon and phosphorus source. Afterwards, only two candidate clones were selected for subsequent analysis (34) and two enzymes were identified. The analysis of their predicted catalytic signatures showed that one of them (PhyX) shares a similar motif with the HAPhys. The other product was indicated as a totally new type of phytase (SFPhy) which is encoded by multiple open reading frames but no further analysis or characterization of the enzyme was performed (34).

Few years later, Farias et al. (35) published their findings regarding a β -propeller phytase derived from the functional screening of a metagenomic library cloned in the pWEB-TNC™ cosmid. The metagenomic DNA was originally obtained from red rice crop plants and castor bean cake. In contrast to Tan and colleagues, Farias et al. did not report multiple positive clones. Only one positive clone was described and its product (PhyRC) purified and comprehensively characterized (35).

Similar to Tan et al. (34), we also confirmed the existence of new types of phytate degrading enzymes not related to the currently accepted groups of phytate degrading enzymes (chapters 3 and 5). The products Pho07, Mblp01 and Mblp02 are remarkable examples of non-classical phytases. The proteins harboring the ALP_like domain like Pho07 are known as promiscuous enzymes that catalyze phosphoryl transfer reactions involved in key processes including energy storage, biosynthesis, or replication of genetic material. This type of enzymes include phosphomono and diesterases, phosphoglycerate mutases, phosphopentomutases, and sulfatases (36) but phytate was not part of the degradable substrates reported for this enzyme types previously. Although, the β -propeller phytases have been reported as alkaline phytate-degrading enzymes, they are not phylogenetically or biochemically related with the enzymes harboring the ALP_like domain(37, 38).

Mlbp01 and Mlp02 belong to Metallo- β -lactamases (MBLs) which have never been reported comprising phytate-degrading representatives. MBLs are considered to be one

of the most promiscuous groups of enzymes. Representatives of the MBLs are capable of catalyzing a variety of different reactions by acting on numerous dissimilar substrates (39). In addition, the two different recombinant host strains carrying constructs with the genes *mblp01* and *mblp02* were less susceptible to the action of β -lactam antibiotics. This result suggested catalytic promiscuity in Mblp01 and Mblp02, as in addition to phytate both enzymes are presumably acting on β -lactam antibiotics.

With the discovery of Pho07, Mblp01 and Mblp02, and the previous findings of Tan et al. (34), the potential of function-based metagenomics approaches to retrieve “real new” phytate-degrading enzymes with novel characteristics has been demonstrated. Taking into account the phytate degrading capabilities of Mblp01, Mblp02 and Pho07, this study suggests the existence of two new types of phytate degrading enzymes. The **ALP_likePhys** represented by Pho07 and tightly related to ALP_like phosphatases and sulfatases and the **MBLPhys** associated to MBLs (Fig. 3).

As expected, the application of function-based screenings enables the retrieval of new types of phytate degrading enzymes. However, the published studies (Tan et al., Farias et al, and this study in the chapters 3 and 5) show differences in the amount and types of recovered enzymes (34, 35). The differences are most likely due to variations in the screening methods. One of the differences is the selected vector for library construction. In this study the library was constructed by using a plasmid vector while Tan et al (34) and Farias et al. (35) used fosmids and cosmids respectively. Although we found high variation in the number of positive clones among the screened libraries, we recovered in every case at least one positive clone from the six different soil samples.

The strategy for the selection of positive clones also varied among the studies. The previous studies based their clone selection on the differential size of the growing colonies, assuming that clones carrying phytate-degrading enzymes will grow faster than those with only background enzymatic activities. (34). As described in chapter 2 this study used a simple modified Sperber minimal media supplemented with BCIP to develop a colorimetric screening strategy.

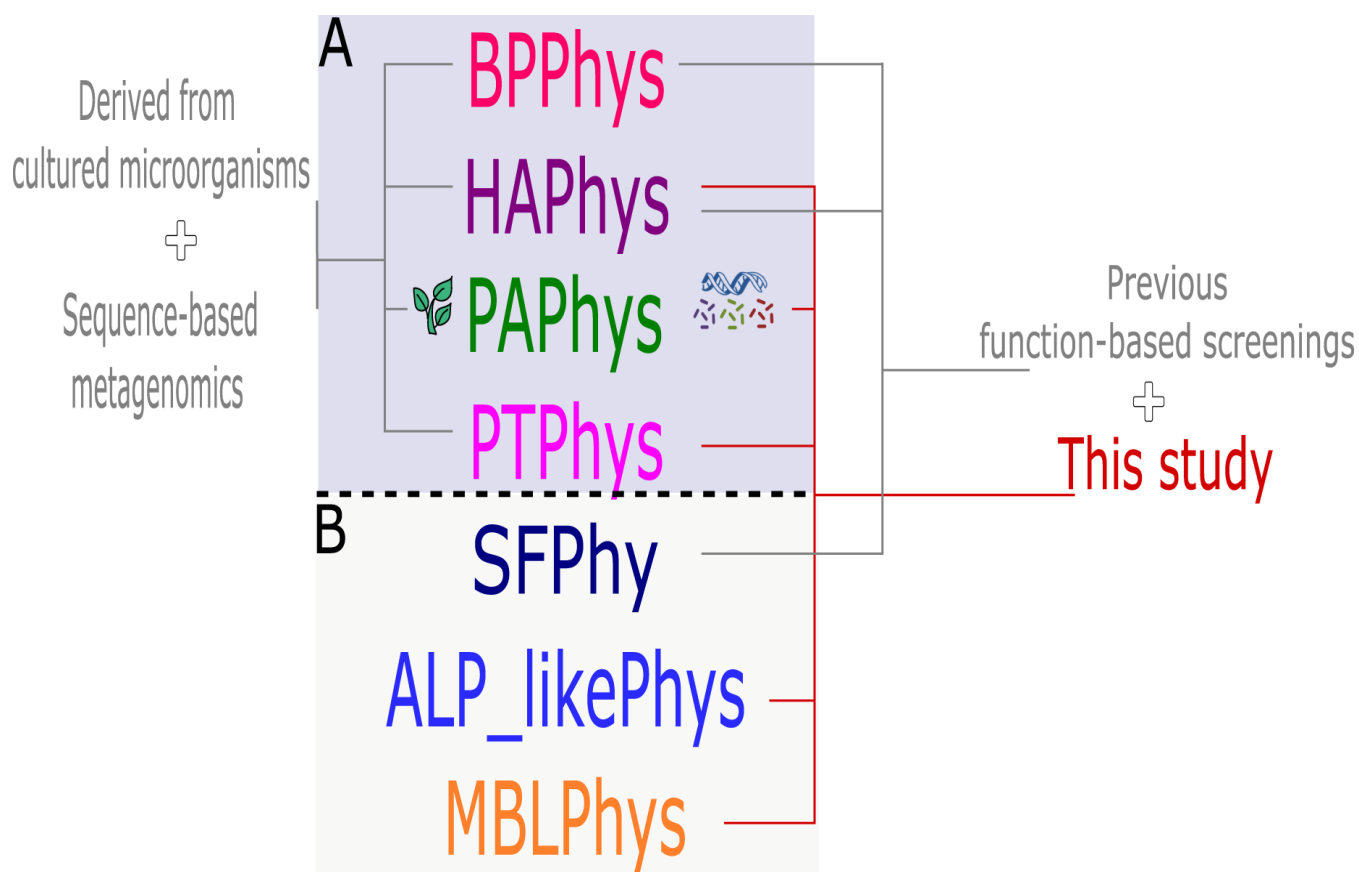


Fig. 3. Different types of phytate degrading enzymes according to their experimental origin. A). Currently accepted types of phytate degrading enzymes derived from culturable microorganisms and sequence-based metagenomics. B). Currently accepted and new types of phytate degrading enzymes derived from this study and previous function-based approaches.

Probably the most significant difference between the previous reports and this study, is the selection of the screening media. Tan et al. (34) and Farias et al. (35) used identical media for screening and selection. Both studies rely on the use of the media described by Unno et al. 2005 (40). That study used two variations of media (solid and semi-solid) supplemented with mineral solutions using phytate as only source of phosphorus and carbon (40). In the present study, the carbon source was not a limiting factor as glucose was included in the indicator medium. Thus, we screened solely for the ability of the library-bearing clones to use phytate as P source. Our method is based on the results of Kerovuoto et al. 1998, who demonstrated, that regardless of the presence of phytate, that complex media and commonly used media containing inorganic phosphates do not induce the expression of phytase-activity in *Bacillus subtilis* (41). In contrast, phytase activity is induced when inorganic phosphate is a limiting factor and phytate is present. This type of phytate/phosphate-mediated regulation is still not fully understood. It is very likely that the previously reported screening approaches are also subject to P availability-dependent

regulation of phytase production but the role of carbon source limitation on expression has not been studied. The use of phytate as sole source of phosphorus and carbon might have an impact on the number of positive clones. Under phosphorus and carbon limitation the enzymes should be able to not only release phosphate, but also to process the *myo*-inositol ring in order to fulfill the carbon demands. This strategy could increase the probabilities of detecting enzymes highly specific for phytate degradation, but limits the screening potential to detect a broader range of phosphatases removing mainly phosphate from phytate. In the present study more phosphatases with proven and indicated phytate-degrading activity than in any other function-based metagenomic survey were recovered. The new phosphatases are broad range substrate enzymes.

Interestingly, this study recovered representatives of several protein families which are known for their broad substrate spectrum e.g. ALP_like, MBLs and the HADs. Enzymes with substrate ambiguity have been gaining importance, because they can provide benefits to cells by different mechanisms. Some of these enzymes are involved in processes such as proofreading, nutrients scavenging, removal of antimetabolites, balancing of metabolite pools or establishing system redundancy (42). Such characteristics might be interesting for industrial applications. Furthermore, the study of substrate ambiguity provides novel information regarding the evolution of different enzymatic clusters as well as the role of the enzymatic promiscuity in element cycles.

6.5. *The Metagenome Input to Phytase Research*

Phytases most known application is the utilization as feed additive for monogastric animals used for large scale production. Multiple alternatives to improve certain enzymatic characteristics in the existent phytases have been implemented with the aim of obtaining an “ideal phytase”. Notwithstanding, with the current state of the art, it is a challenging task to fulfill all the requirements for such an endeavor. The requirements include fast degradability in the intestines and high efficiency to process phytate under the conditions of the digestive tract of the animals. Moreover, it must show resilience to the high temperatures during pelleting process of food manufacture. All accompanied by low production costs (43, 44). In some countries i.e. U.S the production of “ideal” phytases is researched by using genetically engineered enzymes and strains (16, 45). Currently, phytase research does not receive much input from metagenomics in order to improve

and engineer phytases with better characteristics compared to other types of commercially relevant biocatalyst, e.g. lipases (46, 47). The number of lipases and families reported during the last years has increased due to the implementation of large-scale metagenome-based surveys (32, 48, 49). Our results revealed an overlooked diversity of phytate-degrading enzymes combined with broad range of different features of the recovered enzymes. Optimal pH values varied from acid to alkaline, and a broad range of responses to additives such as inhibitors and cations was also observed.

6.6. Function-Based Metagenomics Reveals Classic Phytases with Special Characteristics and New Phosphatases

Phytases importance is not exclusively based on their involvement in animal production or phytate recycling for environmental decontamination. Microorganisms also need to fulfill their own phosphate needs accessing phytate in the environment through extracellular phytases. Microorganisms without extracellular phytases can incorporate phytic acid into cells via inositol phosphate transporters, for further P processing (37, 41). The biological functions of phytate-degrading enzymes are not limited to the retrieval of phosphorus from phytate for nutritional purposes. This type of enzymes is also involved in cell signaling activities, germination in plants and directly or indirectly related with bacterial pathogenicity (37, 50). Several studies showed that the expression of phytases seem to be involved in the modulation of pathways related to the immune response or increase the virulence of certain pathogens (50, 51). Table 1 shows several examples of phytate-degrading activities related to bacterial pathogenicity. In the present study, we found representatives of the classic phytases. One of those representatives Pho16B chapter 4, belongs to the PTP_{Phy}, but it possesses an uncommon catalytic sub-loop (HCXXGKDR[TA]G). This catalytic signature is homologous to the sub-loop present in the atypical lipid phosphatases (ALPs). ALPs are a type of PTPs considered as important virulence factors of human pathogens i.e. *Mycobacterium tuberculosis* (52, 53). Pho16B is the first tyrosine phytase derived from a soil metagenome, and provides a hint on the occurrence and functionality of pathogen-related catalytic signatures in environmental samples. But the nature of our study did not allow to directly show the link between phytases and virulent PTPs/ALPs. Most of the information on the occurrence and evolution of environmentally derived genes involved in pathogenicity is limited to

sequence homology studies, and very little is published on functional genes carrying this type of molecular signatures.

Table 1. Examples of phytase activity and its relationship with diseases in different species. N.D: Not described

Strain/Enzyme/ ORF	Phytase type	Putative affected organism	Source	Effect	Ref
XopH	PTPhy	Plants	<i>Xanthomonas</i>	Manipulate signaling pathways to the pathogen's benefit	(50)
PhyA	BPPhy	Plants	<i>Xanthomonas</i>	Improve virulence	(54)
LppA	PTPhy	Mammals	<i>Legionella pneumophila</i>	Counteracts phytate action which normally restrict intracellular bacterial growth	(55)
SopB	N.D	Humans	<i>Salmonella typhimurium</i>	Effector for bacterial entry	(56)
orf19.3727	N.D	Humans	<i>Candida albicans</i>	Increase virulence potency	(57)
<i>Candida</i> strains	N.D	Humans	<i>Candida glabrata</i>	Pathogen survival and persistence in the host	(58)
<i>Pseudomonas sp.</i> , <i>Psychrobacter sp</i> Strains	N.D	Fish	<i>Pseudomonas sp.</i> , <i>Psychrobacter sp</i>	Stimulates intracellular head kidney leukocyte activities	(59)

One more representative of the classic phytases obtained during the metagenomic screening was the enzyme Pho18 (chapter 3). Pho18 as part of the insert pLP18 was taxonomically associated to Verrucomicrobia. This enzyme is part of the PAPHys, which before this study were assumed as restricted to plants. Several homologous genes of PAPHys in bacteria and fungi have been reported. However, the expression, purification and characterization of an environmental PAPHy putatively originated from bacteria has not been described. Recently, Nasrabadi et al. 2018 claimed the report of the first bacterial purple acid phytase (60). However, what Nasrabadi et al. (60) did in their study, was to obtain and characterize the phytase activity from bacterial isolates and link that

information with the amplification of a homologous gene from another species with reported acid phosphatases. No gene or protein sequence was reported. Nevertheless, their work together with our study represent the final evidence of PAPHy activity expressed by bacterial strains. This represents also an important change in phytase research, as all types of currently recognized phytases are now also reported to be expressed in bacteria.

Finally, the implementation of the function-based approach also allowed the identification of two proteins with very particular characteristics. The ORFs *pho04* and *pho15* encode two enzymes, which show low levels of sequence identity with alkaline phosphatases. Deeper analysis of the sequences evidenced the lack of domains related to phosphatase activity. The only domain of the products Pho04 and Pho15 is the SNARE-associated domain also known as DedA (61). This type of domain has not been related to any type of catalytic activity. DedA is a highly conserved protein family with representatives in most known genome sequences encoding membrane proteins. Although, in most cases the functionality of this type of protein remains unknown, some references connect this family with roles in cellular processes such as homeostasis, cell division, envelope-related stress responses and loss of proton motive force. Interestingly, in *E coli* the mutations of the DedA family genes *yghB* and *yqjA* produce altered membrane phospholipid levels (62). Our results indicate phosphatase activity in the proteins carrying the SNARE domain, which open an entirely new avenue to study this type of proteins.

With the development and implementation of our method we have expanded the knowledge about phosphatase and phytate-degrading activity in environmental soil samples and communities. By applying the function-based methods now it is possible to explore and unlock phosphatases/phytases in the most complex environments.

6.7. References

1. Nacke H, Goldmann K, Schöning I, Pfeiffer B, Kaiser K, Castillo-Villamizar GA, Schruppf M, Buscot F, Daniel R, Wubet T. 2016. Fine spatial scale variation of soil microbial communities under european beech and norway spruce. *Front Microbiol* 7:2067.
2. Daniel R. 2005. The metagenomics of soil. *Nat Rev Microbiol* 3:470.
3. Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J, Pfeiffer S, Prati D, Renner S, Schöning I, Schumacher U, Wells K, Buscot F, Kalko EKV, Linsenmair KE, Schulze E-D, Weisser WW. 2010. Implementing large-scale and long-term functional biodiversity research: The biodiversity exploratories. *Basic Appl Ecol* 11:473-485.
4. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schruppf M, Daniel R. 2016. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci Rep* 6:33696.
5. Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schruppf M, Daniel R. 2011. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* 6:e17000.
6. Fernández-Arrojo L, Guazzaroni M-E, López-Cortés N, Beloqui A, Ferrer M. 2010. Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol* 21:725-733.
7. Ferrer M, Martínez-Martínez M, Bargiela R, Streit WR, Golyshina OV, Golyshin PN. 2016. Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends. *Microb Biotechnol* 9:22-34.
8. Berini F, Casciello C, Marcone GL, Marinelli F. 2017. Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiol Lett* 364:fnx211.
9. Aziz RK, Breitbart M, Edwards RA. 2010. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* 38:4207-4217.
10. Villamizar GAC, Daniel R, Poehlein A. 2017. First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic *Sporomusa sphaeroides* Strain E (DSM 2875). *Genome Announc* 5: e00037-00017.
11. Castillo Villamizar GA, Poehlein A. 2016. Genome sequence of the acetogenic bacterium *Moorella mulderi* dsm 14980t. *Genome Announc* 4:e00444-00416.
12. Menezes-Blackburn D, Romero D, De La Luz Mora M, Jorquera MA, Crowley DE, Marschner P, Greiner R, Fernández MT. 2011. Identification of β -propeller phytase-encoding genes in culturable *Paenibacillus* and *Bacillus* spp. from the rhizosphere of pasture plants on volcanic soils. *FEMS Microbiol Ecol* 75:163-172.
13. Lim BL, Yeung P, Cheng C, Hill JE. 2007. Distribution and diversity of phytate-mineralizing bacteria. *ISME J* 1:321-330.
14. Mustelin T. 2007. A Brief Introduction to the protein phosphatase families, p 9-22. *In* Moorhead G (ed), *Protein Phosphatase Protocols*. Springer New York, Totowa, NJ, USA.
15. Bhavsar K, Khire JM. 2014. Current research and future perspectives of phytase bioprocessing. *Rsc Adv* 4:26677-26691.
16. Li S, Yang X, Yang S, Zhu M, Wang X. 2012. Technology prospecting on enzymes: application, marketing and engineering. *Comput Struct Biotechnol J* 2:e201209017-e201209017.
17. Dvořáková J. 1998. Phytase: Sources, preparation and exploitation. *Folia Microbiol* 43:323-338.
18. Huang H, Shi P, Wang Y, Luo H, Shao N, Wang G, Yang P, Yao B. 2009. Diversity of beta-propeller phytase genes in the intestinal contents of grass carp provides insight into the release of major phosphorus from phytate in nature. *Appl Environ Microbiol* 75:1508-1516.
19. van Hartingsveldt W, van Zeijl CMJ, Harteveld GM, Gouka RJ, Suykerbuyk MEG, Luiten RGM, van Paridon PA, Selten GCM, Veenstra AE, van Gorcom RFM, van den Hondel CAMJJ. 1993. Cloning, characterization and overexpression of the phytase-encoding gene (*phyA*) of *Aspergillus niger*. *Gene* 127:87-94.

20. Tye A, Siu F, Leung T, Lim B. 2002. Molecular cloning and the biochemical characterization of two novel phytases from *B. subtilis* 168 and *B. licheniformis*. *Appl Microbiol Biotechnol* 59:190-197.
21. Simon C, Daniel R. 2011. Metagenomic analyses: past and future trends. *Appl Environ Microbiol* 77:1153-1161.
22. Iwai S, Chai B, Sul WJ, Cole JR, Hashsham SA, Tiedje JM. 2010. Gene-targeted-metagenomics reveals extensive diversity of aromatic dioxygenase genes in the environment. *The ISME journal* 4:279-285.
23. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35:833.
24. Tan H, Wu X, Xie L, Huang Z, Peng W, Gan B. 2016. Identification and characterization of a mesophilic phytase highly resilient to high-temperatures from a fungus-garden associated metagenome. *Appl Microbiol Biotechnol* 100:2225-2241.
25. Tan H, Wu X, Xie L, Huang Z, Peng W, Gan B. 2016. A novel phytase derived from an acidic peat-soil microbiome showing high stability under acidic plus pepsin conditions. *J Mol Microbiol Biotechnol* 26:291-301.
26. Mootapally CS, Nathani NM, Patel AK, Jakhesara SJ, Joshi CG. 2016. Mining of Ruminant Microbial Phytase (RPHY1) from Metagenomic Data of Mehsani Buffalo Breed: Identification, Gene Cloning, and Characterization. *J Mol Microbiol Biotechnol* 26:252-260.
27. Ngara TR, Zhang H. 2018. Recent Advances in Function-based Metagenomic Screening. *Genomics, Proteomics & Bioinformatics* doi:<https://doi.org/10.1016/j.gpb.2018.01.002>.
28. Kanaya E, Sakabe T, Nguyen NT, Koikeda S, Koga Y, Takano K, Kanaya S. 2010. Cloning of the RNase H genes from a metagenomic DNA library: identification of a new type 1 RNase H without a typical active-site motif. *J Appl Microbiol* 109:974-983.
29. Simon C, Herath J, Rockstroh S, Daniel R. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* 75:2964-2968.
30. Gao W, Wu K, Chen L, Fan H, Zhao Z, Gao B, Wang H, Wei D. 2016. A novel esterase from a marine mud metagenomic library for biocatalytic synthesis of short-chain flavor esters. *Microb Cell Fact* 15:41-41.
31. Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R. 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnol Lett* 34:663-675.
32. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. *FEMS Microbiol Ecol* 78:188-201.
33. Pushpam PL, Rajesh T, Gunasekaran P. 2011. Identification and characterization of alkaline serine protease from goat skin surface metagenome. *AMB Express* 1:3.
34. Tan H, Mooij MJ, Barret M, Hegarty PM, Harington C, Dobson AD, O'Gara F. 2014. Identification of novel phytase genes from an agricultural soil-derived metagenome. *J Microbiol Biotechnol* 24:113-118.
35. Farias N, Almeida I, Meneses C. 2018. New bacterial phytase through metagenomic prospection. *Molecules* 23:448.
36. López-Canut V, Roca M, Bertrán J, Moliner V, Tuñón I. 2011. Promiscuity in alkaline phosphatase superfamily. Unraveling evolution through molecular simulations. *J Am Chem Soc* 133:12050-12062.
37. Greiner R. 2007. Phytate-degrading enzymes: Regulation of synthesis in microorganisms and plants, p 78-96. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.
38. Mullaney E, Ullah AH. 2007. Phytases: Attributes, catalytic mechanisms and applications, p 97-110. *In* Turner BL, Richardson AE, Mullaney EJ (ed), *Inositol phosphates: Linking agriculture and the environment*. CABI, Wallingford, UK.

39. Baier F, Tokuriki N. 2014. Connectivity between catalytic landscapes of the metallo-beta-lactamase superfamily. *J Mol Biol* 426:2442-2456.
40. Unno Y, Okubo K, Wasaki J, Shinano T, Osaki M. 2005. Plant growth promotion abilities and microscale bacterial dynamics in the rhizosphere of Lupin analysed by phytate utilization ability. *Environ Microbiol* 7:396-404.
41. Kerovuo J, Lauraeus M, Nurminen P, Kalkkinen N, Apajalahti J. 1998. Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Appl Environ Microbiol* 64:2079-2085.
42. Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. 2014. Enzyme promiscuity: Engine of evolutionary innovation. *J Biol Chem* 289:30229-30236.
43. Dersjant-Li Y, Awati A, Schulze H, Partridge G. 2015. Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. *J Sci Food Agric* 95:878-896.
44. Garrett JB, Kretz KA, O'Donoghue E, Kerovuo J, Kim W, Barton NR, Hazlewood GP, Short JM, Robertson DE, Gray KA. 2004. Enhancing the thermal tolerance and gastric performance of a microbial phytase for use as a phosphate-mobilizing monogastric-feed supplement. *Appl Environ Microbiol* 70:3041-3046.
45. Chen C-C, Cheng K-J, Ko T-P, Guo R-T. 2015. Current progresses in phytase research: three-dimensional structure and protein engineering. *ChemBioEng Reviews* 2:76-86.
46. López-López O, Cerdán ME, González Siso MI. 2014. New extremophilic lipases and esterases from metagenomics. *Curr Protein Pept Sci* 15:445-455.
47. Sharma PK, Kumar R, Kumar R, Mohammad O, Singh R, Kaur J. 2012. Engineering of a metagenome derived lipase toward thermal tolerance: Effect of asparagine to lysine mutation on the protein surface. *Gene* 491:264-271.
48. Lee MH, Oh KH, Kang CH, Kim JH, Oh TK, Ryu CM, Yoon JH. 2012. Novel metagenome-derived, cold-adapted alkaline phospholipase with superior lipase activity as an intermediate between phospholipase and lipase. *Appl Environ Microbiol* 78:4959-4966.
49. Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L. 2008. Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* 133:42-49.
50. Blüher D, Laha D, Thieme S, Hofer A, Eschen-Lippold L, Masch A, Balcke G, Pavlovic I, Nagel O, Schonsky A, Hinkelmann R, Worner J, Parvin N, Greiner R, Weber S, Tissier A, Schutkowski M, Lee J, Jessen H, Schaaf G, Bonas U. 2017. A 1-phytase type III effector interferes with plant hormone signaling. *Nat Commun* 8:2159.
51. Cottrill MA, Golovan SP, Phillips JP, Forsberg CW. 2002. Inositol phosphatase activity of the *Escherichia coli* agp-encoded acid glucose-1-phosphatase. *Can J Microbiol* 48:801-809.
52. Beresford N, Patel S, Armstrong J, Szöör B, Fordham-Skelton Anthony P, Taberero L. 2007. MptpB, a virulence factor from *Mycobacterium tuberculosis*, exhibits triple-specificity phosphatase activity. *Biochem J* 406:13-18.
53. Beresford NJ, Saville C, Bennett HJ, Roberts IS, Taberero L. 2010. A new family of phosphoinositide phosphatases in microorganisms: identification and biochemical analysis. *BMC Genomics* 11:457.
54. Chatterjee S, Sankaranarayanan R, Sonti RV. 2003. PhyA, a secreted protein of *Xanthomonas oryzae* pv. *oryzae*, is required for optimum virulence and growth on phytic acid as a sole phosphate source. *Mol Plant Microbe Interact* 16:973-982.
55. Weber S, Stirnimann CU, Wieser M, Frey D, Meier R, Engelhardt S, Li X, Capitani G, Kammerer RA, Hilbi H. 2014. A type IV translocated Legionella cysteine phytase counteracts intracellular growth restriction by phytate. *J Biol Chem* 289:34175-34188.
56. Zhou D, Chen L-M, Hernandez L, Shears SB, Galán JE. 2001. A Salmonella inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization. *Mol Microbiol* 39:248-260.
57. Tsang PW, Fong WP, Samaranyake LP. 2017. *Candida albicans* orf19.3727 encodes phytase activity and is essential for human tissue damage. *PLoS One* 12:e0189219.

58. Figueiredo-Carvalho MHG, Ramos LdS, Barbedo LS, de Oliveira JCA, Dos Santos ALS, Almeida-Paes R, Zancopé-Oliveira RM. 2017. Relationship between the antifungal susceptibility profile and the production of virulence-related hydrolytic enzymes in brazilian clinical strains of *Candida glabrata*. *Mediators Inflamm* 2017:8952878-8952878.
59. Lazado CC, Caipang CMA, Gallage S, Brinchmann MF, Kiron V. 2010. Responses of Atlantic cod *Gadus morhua* head kidney leukocytes to phytase produced by gastrointestinal-derived bacteria. *Fish Physiol Biochem* 36:883-891.
60. Ghorbani Nasrabadi R, Greiner R, Yamchi A, Nourzadeh Roshan E. 2018. A novel purple acid phytase from an earthworm cast bacterium. *J Sci Food Agric* 98:3667-3674.
61. Doerrler WT, Sikdar R, Kumar S, Boughner LA. 2013. New functions for the ancient DedA membrane protein family. *J Bacteriol* 195:3-11.
62. Thompkins K, Chattopadhyay B, Xiao Y, Henk MC, Doerrler WT. 2008. Temperature sensitivity and cell division defects in an *Escherichia coli* strain with mutations in *yghB* and *yqjA*, encoding related and conserved inner membrane proteins. *J Bacteriol* 190:4489-4500.

7 Summary

Phosphatases, including phytases, play a major role in phosphorous cycle, cell metabolism, biotechnology, and pathogenic processes. Despite their relevance as biocatalysts, the study of phosphatases/phytases has been almost exclusively based on genes and enzymes derived from individual organisms, leaving out the ones associated to microbial communities and so far, uncultured microorganisms. One of the main limitations for the application of function-based metagenomic analysis on phosphatases and particularly phytases, is the lack of reliable and efficient screening methods.

In this study we developed and standardized a function-based method to retrieve novel phosphatases/phytases from soil samples. The developed method is based on the induction of phosphatase/phytase activity in metagenomic library-bearing *E. coli* strains by the absence of inorganic phosphorus sources and the presence of phytate as sole phosphorus source in minimal media. A total of 612,910 clones from five different metagenomic soil libraries were analyzed. Twenty-three positive *E. coli* clones were recovered and their respective inserts sequenced and analyzed. The phylogenetic analysis of the inserts indicated that all cloned environmental DNA fragments were of bacterial origin. Sequence similarity searches of the 23 inserts resulted in the identification of 33 candidate genes. The deduced gene products of 31 of these genes showed at least some similarity to known phosphatases whereas the remaining two were similar to metallo-beta-lactamases. The protein sequence analysis revealed the presence of 11 different domains in the 33 deduced proteins. We detected the alkaline phosphatase and sulfatase superfamily domain (ALP-like cl23718) as the most frequent domain with eight representatives. In addition, two phosphatase-related genes of plasmids pLP04 and pLP15 did not encode known catalytic domains or signatures directly or indirectly associated with phosphatases. The produced proteins Pho4 and Pho15, respectively, showed phosphatase activity and low sequence identity to alkaline phosphatases, but the only catalytic domain detected in both proteins was the SNARE-associated superfamily domain (cl00429).

In total, five genes out of the 33 candidates were selected on the basis of their biological significance for the individual subcloning, expression and characterization of the respective products. Pho07 was selected as representative of the most abundant superfamily detected in our survey from which to our knowledge no enzyme has been

reported or comprehensively characterized regarding phytase activity. Pho18 was selected as a member of the classic purple acid phytases, which before this study were almost exclusively represented by plant-derived phytases. Pho16B was analyzed in more detail due to its affiliation to the protein tyrosine phytases. In addition, Pho16B possesses a molecular signature similar to members of a tyrosine phosphatases subtype, which are closely related to microbial pathogenesis. Finally, Mblp01 and Mblp02 were selected, as they are representatives of the metallo- β -lactamases. This superfamily is one of the most promiscuous types of enzymes. Moreover, the metallo- β -lactamases share an evolutionary relationship with phosphatases but did not comprise representatives with reported phytase activity prior to this study.

The characterization of the selected proteins revealed different features among all analyzed enzymes. The lowest optimal pH of enzyme activity was 4.0 (Pho07) and the highest 7.0 (Mblp02). The highest optimal temperature was recorded at 50 °C (Pho18 and Mblp02) and the lowest at 30 °C (Pho07). The substrate specificity, was also variable among the analyzed enzymes. All 5 purified enzymes were able to degrade phytate but phytate was the preferred substrate only for Pho07. The metallo- β -lactamases (Mblp01 and Mblp02) showed their highest activity with ATP. Naphthyl phosphate and pyrophosphate were the preferred substrates for Pho16B and Pho18, respectively. Interestingly, both metallo- β -lactamases show potential mechanistic promiscuity as the ability to act on β -lactam antibiotics is indicated by reduced sensitivity of recombinant strains harboring *mblp01* or *mblp02* towards β -lactam antibiotics.

This study provided a reliable function-based screening method to recover novel phosphatases/phytases and insights into the diversity of enzymes capable of phytate degradation. This study also demonstrates the potential of functional metagenomics to exploit phosphatase/phytase pools hidden in environmental samples by recovering novel phytate-degrading enzymes unrelated to previously known types of phytases.

8 Appendix

8.1. Thesis Declaration

Declaration of plagiarism

I hereby confirm that I have written this doctoral thesis independently. I have not used other sources or facilities others than the ones mentioned in the chapters. The contributions of the authors are given proceeding the respective manuscripts. Moreover, I have not used unauthorized assistance and have not submitted this thesis previously in any form for another degree at any institution or university

Genis Andrés Castillo Villamizar

8.2. Additional Publications

8.2.1.

Fine Spatial Scale Variation of Soil Microbial Communities under European Beech and Norway Spruce

Heiko Nacke¹, Kezia Goldmann^{2,3}, Ingo Schöning⁴, Birgit Pfeiffer¹, Kristin Kaiser¹, **Genis Andrés Castillo-Villamizar**¹, Marion Schrumpf⁴, François Buscot^{2,5}, Rolf Daniel¹, Tesfaye Wubet^{2,5}

Front Microbiol (2016) 7:2067

¹*Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany,*

²*Department of Soil Ecology, UFZ-Helmholtz Centre for Environmental Research, Halle, Germany,* ³*Department of Biology II, University of Leipzig, Leipzig, Germany,* ⁴*Max Planck Institute for Biogeochemistry, Jena, Germany,* ⁵*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany*

Author contributions to the work:

Performed the experiments: HN, KG, IS, BP, KK, and GC.

Analyzed data: HN, KG, IS, and KK.

Wrote the paper: HN, KG and IS

Conceived and designed the experiments: MS, FB, RD, and TW



Fine Spatial Scale Variation of Soil Microbial Communities under European Beech and Norway Spruce

Heiko Nacke^{1†*}, Kezia Goldmann^{2,3†}, Ingo Schöning^{4†}, Birgit Pfeiffer¹, Kristin Kaiser¹, Genis A. Castillo-Villamizar¹, Marion Schrupf⁴, François Buscot^{2,5}, Rolf Daniel¹ and Tesfaye Wubet^{2,5}

¹ Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen, Germany, ² Department of Soil Ecology, UFZ-Helmholtz Centre for Environmental Research, Halle, Germany, ³ Department of Biology II, University of Leipzig, Leipzig, Germany, ⁴ Max Planck Institute for Biogeochemistry, Jena, Germany, ⁵ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

OPEN ACCESS

Edited by:

Tim Daniell,
James Hutton Institute, UK

Reviewed by:

Christopher Blackwood,
Kent State University, USA
Richard S. Winder,
Natural Resources Canada, Canada
Christina Hazard,
University of Lyon, France

*Correspondence:

Heiko Nacke
hnacke@gwdg.de

[†]These authors have contributed equally to this work.

Specialty section:

This article was submitted to
Terrestrial Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 15 August 2016

Accepted: 07 December 2016

Published: 22 December 2016

Citation:

Nacke H, Goldmann K, Schöning I, Pfeiffer B, Kaiser K, Castillo-Villamizar GA, Schrupf M, Buscot F, Daniel R and Wubet T (2016) Fine Spatial Scale Variation of Soil Microbial Communities under European Beech and Norway Spruce. *Front. Microbiol.* 7:2067. doi: 10.3389/fmicb.2016.02067

The complex interactions between trees and soil microbes in forests as well as their inherent seasonal and spatial variations are poorly understood. In this study, we analyzed the effects of major European tree species (*Fagus sylvatica* L. and *Picea abies* (L.) Karst) on soil bacterial and fungal communities. Mineral soil samples were collected from different depths (0–10, 10–20 cm) and at different horizontal distances from beech or spruce trunks (0.5, 1.5, 2.5, 3.5 m) in early summer and autumn. We assessed the composition of soil bacterial and fungal communities based on 16S rRNA gene and ITS DNA sequences. Community composition of bacteria and fungi was most strongly affected by soil pH and tree species. Different ectomycorrhizal fungi (e.g., *Tylospora*) known to establish mutualistic associations with plant roots showed a tree species preference. Moreover, bacterial and fungal community composition showed spatial and seasonal shifts in soil surrounding beech and spruce. The relative abundance of saprotrophic fungi was higher at a depth of 0–10 vs. 10–20 cm depth. This was presumably a result of changes in nutrient availability, as litter input and organic carbon content decreased with soil depth. Overall bacterial community composition showed strong variations under spruce with increasing distance from the tree trunks, which might be attributed in part to higher fine root biomass near spruce trunks. Furthermore, overall bacterial community composition was strongly affected by season under deciduous trees.

Keywords: tree species, soil depth, horizontal distance from tree trunk, seasons, soil properties, soil microbial community structure, bacterial 16S rRNA gene, fungal ITS DNA

INTRODUCTION

Earth currently harbors approximately three trillion trees and only one gram of soil can contain billions of microbial cells (Rosselló-Mora and Amann, 2001; Crowther et al., 2015). The effect of trees on bacteria and fungi in forest soils, comprising many taxa involved in decomposition of plant litter as well as deadwood, is however poorly understood (Wubet et al., 2012; Pfeiffer et al., 2013; Purahong et al., 2014). Forest trees substantially impact soil physical, chemical and biological

properties by species-specific stemflow, root architecture, leaf and root litter inputs, root exudates, nutrient uptake, shade, and microclimate (Augusto et al., 2002; Ayres et al., 2009; Raz-Yaseef et al., 2010; Cesarz et al., 2013). As a consequence of direct or indirect tree impacts, changes in the spatial distribution of microbes, vertically through the soil profile as well as horizontally with increasing distance from tree trunks, can occur (Saetre and Bååth, 2000; Ettema and Wardle, 2002). Although numerous studies on the effects of plants on soil microorganisms are available, they rarely focus on microbial communities under trees (Thoms et al., 2010; Urbanová et al., 2015; Uroz et al., 2016). Surveys on effects of pure tree species in a forest stand as well as those focusing on vegetation gradients or chronosequences contributed to the current overall picture concerning tree influences on soil microbial communities (e.g., Cong et al., 2015; Zeng et al., 2016).

European beech (*Fagus sylvatica* L.) and Norway spruce (*Picea abies* (L.) Karst) represent dominant forest trees in Central Europe (Cesarz et al., 2013; Hanewinkel et al., 2013). Since the 19th century, reforestation of devastated forest sites using Norway spruce has been very common in Central Europe (Berger and Berger, 2012). Beech forests show a high seasonal variation in aboveground litter input, which is predominately autumnal. In contrast, the aboveground litter input in spruce forest remains relatively constant over the year. Components of needle litter from Norway spruce such as waxes and phenolic compounds are highly recalcitrant to biological degradation, whereas beech leaf litter contains higher amounts of more readily decomposed water-soluble substances (Nykvist, 1963; Priha and Smolander, 1997). Replacement of beech by spruce species is therefore accompanied by changes in humus form, acidity and soil structure (Berger and Berger, 2012). Upper soil horizons are dominated by leaf litter input, and roots; their residues and exudation patterns shape the subsoil (Moll et al., 2015). Spruce is typically shallow-rooted, whereas beech has a deep rooting system (so called “base-pump”). Consequently, variation in nutrient availability affects microbial communities along soil depths (Huang et al., 2013; Moll et al., 2015). Between *Fagus sylvatica* L. and *Picea abies* (L.) Karst, the quantity and composition of exudates varies with season (Geßler et al., 1998; Fender et al., 2013) and potentially affects microbial processes such as respiration (Cesarz et al., 2013).

European beech and Norway spruce forest stands differ in the magnitude of stemflow. In beech stands, stemflow water contributes 5–20% to the annual soil water input (Koch and Matzner, 1993; Johnson and Lehmann, 2006). Stemflow in conifer forests is often below 1% due to differences in branch angle, specific surface roughness of branches and bark (Johnson and Lehmann, 2006). The high stemflow in beech forests is associated by a decrease of soil pH next to the stem base versus the surrounding soil (Koch and Matzner, 1993). A similar effect has not been demonstrated in Norway spruce forest.

Previous studies have largely used methods providing coarse phylogenetic information to identify effects of forests on soil microbial communities. Using automated ribosomal intergenic spacer analysis (ARISA), ester linked fatty acid methyl ester (EL-FAME) analyses, and denaturing gradient gel electrophoresis

(DGGE), differences in soil bacterial and fungal community structure in temperate broad-leaved and coniferous forests have been reported (Lejon et al., 2005; Zechmeister-Boltenstern et al., 2011; Jiang et al., 2012). Recently, Tedersoo et al. (2016) analyzed pyrosequencing-derived ITS sequences to assess the effects of tree diversity on fungi, protists and meiofauna inhabiting forest soil. Results indicated that compared to the effects of individual tree species and soil parameters, tree diversity *per se* had a minor influence on the taxonomic richness of soil biota (Tedersoo et al., 2016). In addition, based on amplicon pyrosequencing data, significant effects of tree species on soil bacterial and fungal community composition were reported by Urbanová et al. (2015).

While several recent marker gene sequencing-based studies focused either on bacteria or fungi in forest soils, they have rarely been considered together (Yarwood et al., 2010; Baldrian et al., 2012; Urbanová et al., 2015). Fungi are typically larger in size than bacteria and exhibit a higher biomass. Therefore, they interact with their environment, e.g., by moving water and nutrients, on a larger spatial scale compared to bacteria (Coleman and Crossley, 1996; van der Heijden et al., 2008; Trevors, 2010), which might result in a more homogeneous distribution of fungal communities in soil. The life cycle of both bacteria and fungi inhabiting forest soils can be strongly affected by seasons through changes in abiotic and biotic factors (Thoms and Gleixner, 2013).

In this study, we applied pyrosequencing of the V3–V5 region of the 16S rRNA gene and the ITS DNA region to assess composition of soil bacterial and fungal communities in a European beech and a Norway spruce forest. We considered potential seasonal variation in microbial communities by collecting samples in early summer and autumn. Furthermore, to determine spatial tree effects, soil collected from different depths and horizontal distances toward tree trunks was considered within this survey. We examined the following hypotheses: (1) bacterial and fungal community composition are affected by tree species, (2) the relative abundance of saprotrophic microorganisms decreases with soil depth, (3) bacteria respond stronger to growing distance from trees than fungi, and (4) seasonal variation of soil bacterial and fungal community composition is stronger under deciduous versus coniferous forests.

MATERIALS AND METHODS

Sites and Soil Sampling

All soil samples were derived from a beech (*Fagus sylvatica* L.) and a spruce (*Picea abies* L. (Karst)) forest site (distance between the two forest sites: approximately 5 km) located in the Hainich-Dün region in Germany (Fischer et al., 2010). The beech and spruce forest stands were originally established as plantations and are managed (management type, age class forest) since 1760 and approximately 1930, respectively (Wäldchen et al., 2011). Due to the very fertile soils (the original parent material was limestone covered by loess) at both sites, beech-dominated forest would be the natural forest type. The age of the trees at both sites ranged between 50 and 65 y. Beech and spruce trees exhibited average crown radii of 387 ± 29 and 209 ± 12 cm, respectively.

The mineral soil was sampled at 0–10 cm and 10–20 cm depth using a split tube sampler with a diameter of 4.8 cm (Eijkelkamp Agrisearch Equipment, Giesbeek, Netherlands). Mineral soil samples were taken from different horizontal distances (0.5, 1.5, 2.5, and 3.5 m) from the trunks of four randomly-selected trees per site (“tree distance”; see **Figure 1**). Sampling was performed in two seasons, early summer and autumn 2012. Five year averages (2008–2012) of soil temperature, measured at a depth

of 10 cm, showed similar seasonal variations in the beech (May: 12.1°C, November: 4.1°C) and spruce forest stand (May: 12.7°C, November: 4.2°C). We applied a paired sampling. The sampling positions in autumn were <30 cm away from the sampling points in early summer (Table S1). All sampling points showed a distance >3.5 m to tree trunks (except trunks of the four selected beech and spruce trees, respectively). In total 128 soil samples (2 sites × 2 seasons × 4 replicate trees × 4 horizontal distances × 2 soil depths) were immediately sieved to <4 mm in the field and individually homogenized. One subsample (>200 g) of each sample was air-dried and sieved to <2 mm for soil chemical analyses and another subsample (50 g) was frozen (−20°C) for extraction of nucleic acids.

Soil Physical and Chemical Properties

Soil pH was measured in duplicate in the supernatant of 1:2.5 mixtures of soil and aqueous 0.01 M CaCl₂ with a glass electrode. Additionally, the gravimetric water content of the air-dried soil was determined. The empirical equation of Wäldchen et al. (2012) was used to estimate clay content in the samples. The remaining soil was ground to <100 μm. Ground samples were analyzed for total carbon (TC) and nitrogen (TN) by dry combustion with the CN analyzer “Vario Max”™ (Elementar Analysensysteme GmbH, Hanau, Germany). Inorganic carbon (IC) concentrations were determined with the same analyzer after the ignition of samples for 16 h at 450°C. The organic carbon (OC) concentrations equaled the differences between TC and IC.

DNA Extraction, Amplification and Pyrosequencing

Total microbial community DNA was extracted from approximately 2 g of frozen soil per sample using the PowerSoil™ total RNA isolation kit, the PowerSoil™ DNA elution accessory kit, and the PowerClean™ DNA Clean-Up kit (MoBio Laboratories, Carlsbad, CA, USA) according to the instruction. DNA concentrations were quantified using a NanoDrop UV-Vis spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany).

The V3–V5 region of bacterial 16S rRNA genes was amplified by PCR. The following set of primers containing Roche 454 pyrosequencing adaptors (underlined) and a sample-specific MID (Extended Multiplex Identifier) was used: V3for 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAG-MID-TA CGGRAGGCAGCAG-3' (Liu et al., 2007) and V5rev 5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCCGTCAATT CMTTGTGAGT-3' (Wang and Qian, 2009). The PCR reaction mixture (50 μl) contained 10 μl 5-fold reaction buffer (Phusion HF buffer, Thermo Fisher Scientific Inc., Germany), 200 μM of each of the four deoxynucleoside triphosphates, 5% DMSO, 1 U Phusion high fidelity DNA polymerase (Thermo Fisher Scientific Inc.), approximately 25 ng DNA as template, and 4 μM of each of the primers. The PCR reactions were initiated at 98°C (2 min), followed by 25 cycles of 98°C (45 s), 58°C (45 s), and 72°C (40 s), and ended with incubation at 72°C for 5 min.

Fungal ITS DNA was amplified using primer ITS1F (Gardes and Bruns, 1993) containing a sample-specific MID and Roche 454 pyrosequencing adaptor B and primer ITS4 (White et al.,

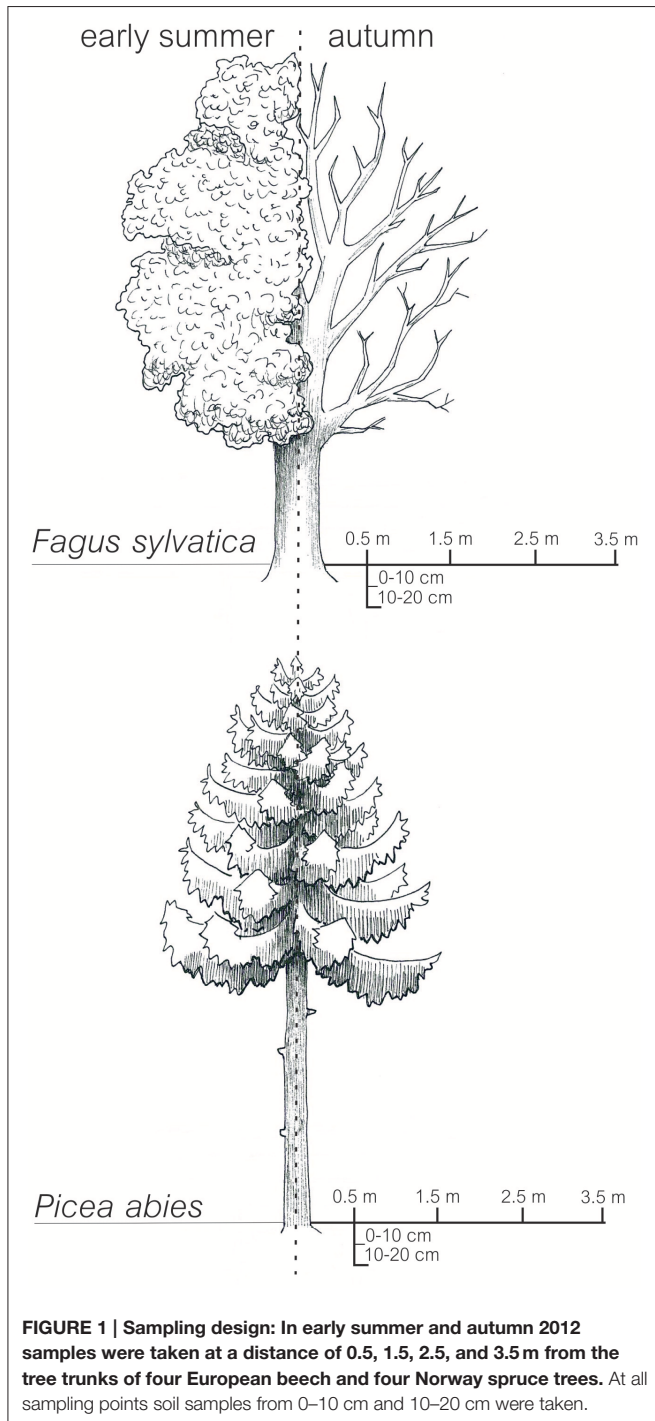


FIGURE 1 | Sampling design: In early summer and autumn 2012 samples were taken at a distance of 0.5, 1.5, 2.5, and 3.5 m from the tree trunks of four European beech and four Norway spruce trees. At all sampling points soil samples from 0–10 cm and 10–20 cm were taken.

1990) containing Roche 454 pyrosequencing adaptor A. The PCR reactions were performed in a total volume of 50 μ l reaction mix containing 1 μ l DNA template (7–15 ng), 25 μ l Go Taq Green Master mix (Promega, Mannheim, Germany) and 1 μ l 25 pmol of each of the ITS region-specific primers. Touchdown PCR conditions as described by Wubet et al. (2012) were used to amplify fungal ITS DNA.

All samples were amplified in triplicate, purified using the peqGold gel extraction kit (Peqlab Biotechnologie GmbH) and the Qiagen gel extraction kit (Qiagen, Hilden, Germany) as recommended by the manufacturer, and pooled in equal amounts. Quantification of PCR products was performed using the Quant-iT dsDNA BR assay kit and a Qubit fluorometer (Life Technologies GmbH, Karlsruhe, Germany). Sequences of partial 16S rRNA genes and fungal ITS DNA were decoded at the Göttingen Genomics Laboratory and the Department of Soil Ecology (UFZ-Helmholtz Centre for Environmental Research, Halle, Germany), respectively, using a Roche GS-FLX 454 pyrosequencer (Roche, Mannheim, Germany) and Titanium chemistry as recommended by the manufacturer.

The 16S rRNA gene and ITS DNA sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under study accession numbers SRP040766 and SRP044665, respectively.

Sequence Analysis

Bacterial 16S rRNA gene sequence datasets were preprocessed as described by Broszat et al. (2014). Briefly, bacterial sequences shorter than 200 bp, as well as those exhibiting low quality values (<25), more than two primer mismatches, or long homopolymers (>8 bp), were removed using QIIME (Caporaso et al., 2010). In addition, the bioinformatics tools cutadapt (Martin, 2011), Uchime (Edgar et al., 2011), and Acacia (Bragg et al., 2012) were used for truncation of remaining primer sequences, removal of potential chimeric sequences, and removal of noise introduced by amplicon pyrosequencing. Uclust (Edgar, 2010), implemented in QIIME (Caporaso et al., 2010), was used to determine bacterial OTUs at a genetic distance of 3%. To taxonomically classify OTUs, partial 16S rRNA gene sequences were compared with the SILVA SSU database release 119 (Pruesse et al., 2007). OTUs classified as chloroplast or mitochondrion and unclassified OTUs (proportion of unclassified OTUs was approximately 0.2%), which were not affiliated to bacteria, were removed from 16S rRNA gene sequence datasets.

Fungal ITS DNA sequence datasets were preprocessed with Mothur (Schloss et al., 2009) as described by Goldmann et al. (2015). In brief, sequences with ambiguous bases, homopolymers and primer differences (>8 bp) as well as MIDNs were removed in a first filtering step. Simultaneously, short reads (<300 bp), sequences with a low quality score (<20) and noisy sequence ends were removed. Samples were checked for chimeric sequences using the UCHIME algorithm (Edgar et al., 2011). Cd-hit (Li and Godzik, 2006) was applied to determine fungal OTUs at 3% genetic distance. To identify fungi and taxonomically classify OTUs, ITS DNA sequences were queried against the UNITE database (Kõljalg et al., 2013) by using the classify.seq command as implemented in MOTHUR (Schloss et al., 2009). All produced OTUs belonged to the kingdom fungi. To improve

the taxonomical resolution, OTUs that had been assigned only down to the family level were subjected to a BLASTn search (e.g., Johnson et al., 2008) against the NCBI GenBank database (Benson et al., 2015). The searches excluded uncultured and environmental sample sequences and only assignments with a query cover >95%, $E < 0.0001$ and sequence identity >97% were considered. Finally, all fungal OTUs identified at the genus level were grouped into ectomycorrhizal, saprotrophic, and other fungi based on literature.

Bacterial and fungal OTUs comprising only one or two sequences (singleton and doubleton OTUs) were removed from the datasets. The number of analyzed sequences per sample can have an effect on the predicted number of OTUs (Morales et al., 2009). Therefore, OTU-based comparisons were performed at the same level of surveying effort (bacteria: 2540 sequences per sample; fungi: 1996 sequences per sample). In this study, we focused on microbial community composition. Data on microbial diversity is provided in the Supplementary Material (see **Figures S1, S2**). OTUs identified at a genetic distance of 3% were used to calculate rarefaction curves and the Shannon index.

Statistical Analyses

The response of main soil characteristics (e.g., C:N ratio, clay content) to soil depth (0–10 and 10–20 cm depth), season (early summer and autumn) and tree distance (0.5, 1.5, 2.5, and 3.5 m) was assessed for both study sites separately by analysis of covariance (ANCOVA) using the “aov” command of the “Stats” R-package (R Development Core Team, 2015). The random effects of the four sampling transects per study site were considered in the analysis by including them as a factor in our linear models (tree replicate).

The effect of tree species on soil bacterial and fungal community composition, respectively, was visualized using principal coordinates analysis plots generated with the emperor software package (Vázquez-Baeza et al., 2013) and the “ordplot” function incorporating environmental vectors calculated with the “envfit” function of the “Vegan” R-package (Oksanen et al., 2016). In order to test the effects of tree replicate, soil pH, OC, soil depth, sampling season, and tree distance on bacterial and fungal community composition, we performed multivariate analysis of variance (MANOVA) using the “adonis” command of the “Vegan” R-package (Oksanen et al., 2016) based on weighted UniFrac (Lozupone et al., 2011) distance matrices. The adonis function in R implements a sequential sum of squares (type 1). A priori we decided to include first the random variance of the tree replicates and important abiotic drivers (soil pH and organic C) into the model. In a second step the factors depth, season and distance were added. This means that the significance of depth, season and distance was examined after removal of variance explained by soil pH and organic C concentration. Changing the order of soil pH and organic C or the order of depth, season and distance in the model would not change the significance of the individual factors. This can be explained by the missing collinearity among these factors. These analyses were conducted for whole microbial communities and microbial communities under each tree species individually. Adjusted R^2 -values of total models increased with the addition of every single considered parameter (**Tables S2, S3**).

To further identify individual taxa strongly associated with a specific tree species, season or spatial position in soil, the multipatt algorithm and the “IndVal” function in the “Indicspecies” R-package (De Cáceres and Legendre, 2009) was used based on bacterial and fungal OTUs. The PAST statistical package (Hammer et al., 2001) was used for the performance of Mann-Whitney U test and Spearman’s rank correlations. We applied Mann-Whitney U test to identify dominant genera showing significant differences between sets of samples. Spearman’s rank correlations were used to correlate relative abundances of dominant genera with soil parameters.

RESULTS

General Characteristics of Soil Samples

Both forest stands grow on limestone, which is covered with a loess layer of variable thickness. The loess layer is thinner at the spruce than at the beech forest site. Therefore, in 0–10 cm depth pH values ranged between 3.1 and 5.9 at the spruce site and between 3.7 and 4.4 at the beech site (Table 1, Table S4). The pH values determined for our samples are typical for the two forest sites. At 5 out of 32 sampling locations within the spruce forest the pH at a depth of 0–10 cm was >5.5 indicating that the loess layer was less pronounced or absent and that the parent material mainly originated from limestone. We did not detect a decrease of the soil pH next to the stem basis of beech trees (0.5 m tree distance) compared to the other considered sampling distances (Table S4). At a depth of 10–20 cm the average pH increased by 0.9 units in the spruce stand, whereas it decreased by 0.2 units in the beech stand, which is again a result of the lower loess layer thickness in the spruce compared to the beech stand. This was confirmed by the clay content (0–10 cm), which was with $388 \pm 15.2 \text{ g kg}^{-1}$ (mean \pm standard error) on average higher at the spruce than at the beech site with $276 \pm 4.4 \text{ g kg}^{-1}$. At the 0–10 cm depth, the soils contained on average $32.6 \pm 2.3 \text{ g kg}^{-1}$ and $26.2 \pm 0.8 \text{ g kg}^{-1}$ OC in the spruce and beech stand. The OC concentrations decreased with depth. Organic C concentrations at the 0- to 10-cm depth were strongly related to estimated clay contents ($r = 0.79$, $P < 0.001$). Due to collinearity between OC concentration, clay content, and C:N ratio, we only included OC concentration in subsequent statistical analyses.

Soil Bacterial and Fungal Community Profiles

Pyrotag processing yielded a total of 864,096 bacterial and 255,488 fungal high-quality sequences with an average length

of 464 and 300 bp, respectively. At a genetic distance of 3%, 23,727 bacterial and 1336 fungal OTUs were identified across all analyzed soil samples. In the final microbial dataset, the number of OTUs per individual soil sample ranged from 505 to 1440 (bacteria) and 45 to 191 (fungi). Taxonomic classification was based on closest matches of OTUs to particular phylogenetic groups. Each of the dominant phyla and genera identified in this study (see Figures 2, 3) is represented by more than one OTU determined at a genetic distance of 3%. The bacterial phyla and proteobacterial classes detected in each of the individual soil samples comprised Acidobacteria (average relative abundance: $40.7 \pm 0.8\%$), Alphaproteobacteria ($20.5 \pm 0.4\%$), Actinobacteria ($9.4 \pm 0.3\%$), Gammaproteobacteria ($5.8 \pm 0.2\%$), Chloroflexi, ($4.8 \pm 0.2\%$), Gemmatimonadetes ($4.4 \pm 0.2\%$), Deltaproteobacteria ($3.8 \pm 0.2\%$), Betaproteobacteria ($3.3 \pm 0.1\%$), Bacteroidetes ($2.1 \pm 0.1\%$) and candidate division WPS-2 ($1.5 \pm 0.1\%$) (Figure 2). Genus level analysis of the bacterial community showed high relative abundances (average relative abundance of each genus >1%) of *Bradyrhizobium* followed by *Acidothermus*, *Gemmatimonas*, *Rhizomicrobium*, and *Reyranella* (Figure 3). Acidobacteria represent the most abundant phylum in our study. Subgroup 2 (average relative abundance: $14.1 \pm 0.6\%$), subgroup 1 ($11.1 \pm 0.5\%$), subgroup 3 ($10.1 \pm 0.3\%$), and subgroup 6 ($2.8\% \pm 0.3\%$) showed the highest average relative abundance among acidobacterial representatives.

The fungal community was dominated by Basidiomycota (average relative abundance: $87.7 \pm 0.7\%$), followed by Ascomycota ($8.9 \pm 0.6\%$), and Zygomycota ($2.5 \pm 0.2\%$) (Figure 2). In total, 89% of all dominant fungal OTUs were assigned to more than 200 fungal genera. The most abundant fungal genera were *Russula* (average relative abundance: $33.3 \pm 2.7\%$), followed by *Inocybe* ($16.8 \pm 1.8\%$), *Hygrophorus* ($6.2 \pm 1.0\%$), *Sebacina* ($5.7 \pm 1.0\%$), and *Thelephora* ($5.6 \pm 1.0\%$) (Figure 3). Functional group assignment of the fungal communities revealed that among the 20 most abundant fungal genera (Figure 3), 16 are known to be ectomycorrhizal (ECM) fungi, whereas the remaining four have a saprotrophic lifestyle (*Cryptococcus*, *Mortierella*, *Leotia*, and *Mycena*).

Tree Species Effects on Microbial Community Composition

Samples collected under beech and spruce tend to cluster separately in principal coordinates analysis plots (Figure 4). The axes of these plots explain less of the variability in fungal community composition (axis 1 = 14%) compared to bacterial

TABLE 1 | Basic properties of soil samples derived from the beech and spruce stands.

Origin	Soil depth	pH	Clay content [g kg^{-1}]	Organic C [g kg^{-1}]	C:N ratio
Beech stand	0–10 cm	4.0 ± 0.0	276 ± 4.4	26.2 ± 0.77	12.0 ± 0.10
Beech stand	10–20 cm	3.8 ± 0.0	249 ± 4.2	14.5 ± 0.55	11.0 ± 0.11
Spruce stand	0–10 cm	4.0 ± 0.2	388 ± 15.2	32.6 ± 2.30	14.8 ± 0.27
Spruce stand	10–20 cm	4.9 ± 0.2	380 ± 14.0	15.1 ± 0.65	11.0 ± 0.20

Mean values and standard errors are provided for pH, clay content, organic C, and C:N ratio.

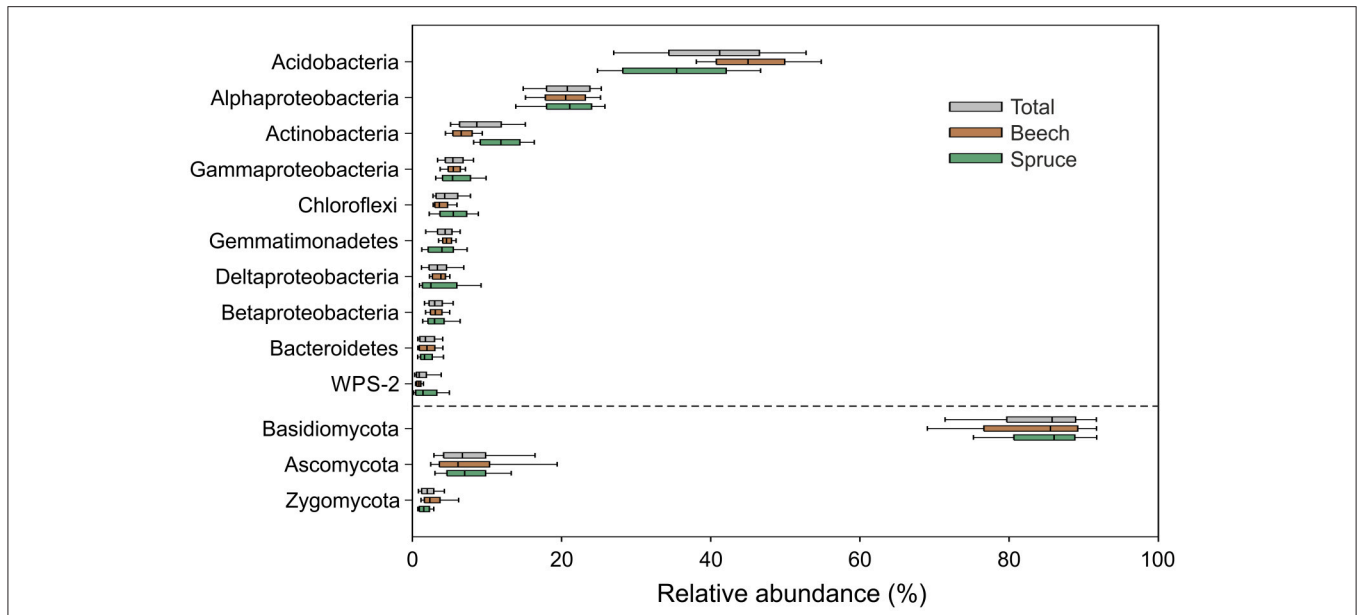


FIGURE 2 | Box-and-whiskers plot showing relative abundances of bacterial and fungal phyla as well as proteobacterial classes detected in each of the analyzed 128 soil samples. Relative abundances of taxa across all samples (gray color) as well as separately with respect to soil surrounding beech (brown color) and spruce (green color) are depicted. The dashed line separates relative abundances of bacterial and fungal taxa.

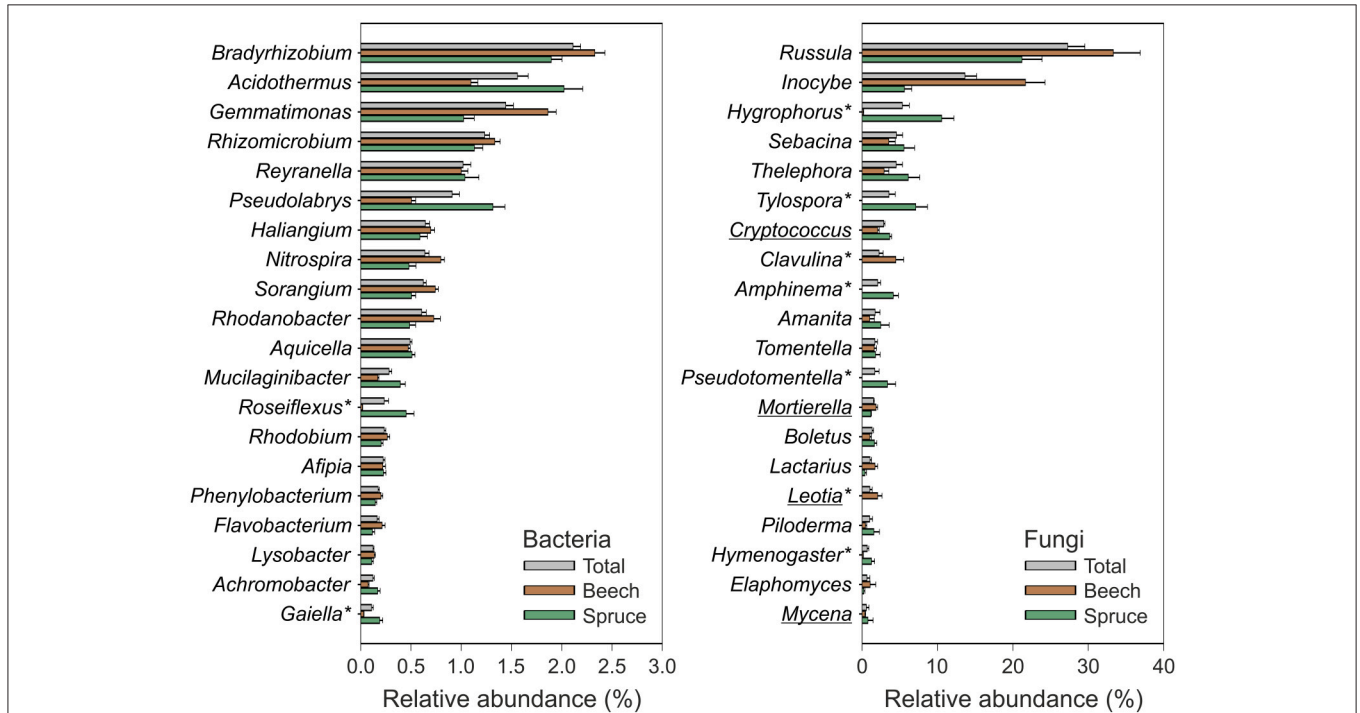
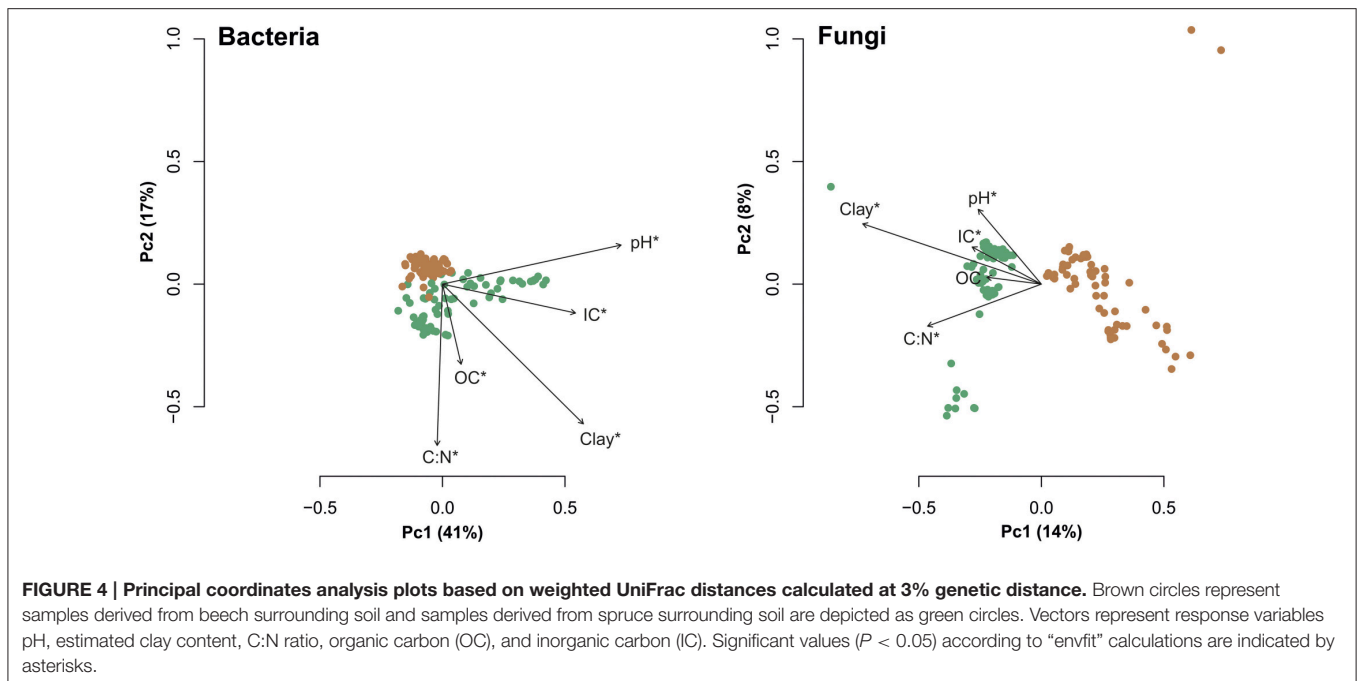


FIGURE 3 | Relative abundance of dominant bacterial and fungal genera detected in the analyzed soil samples. The data represent mean values and standard errors of relative abundance for the 20 most abundant bacterial and fungal genera, respectively. Acidobacteria were analyzed at the subgroup level and therefore not considered within this figure. Relative abundances of taxa across all samples (gray color) as well as separately with respect to soil surrounding beech (brown color) and spruce (green color) are depicted. Asterisks indicate taxa showing an at least five-fold difference in mean relative abundance between spruce and beech ($P < 0.001$ for the Mann-Whitney U test). Underlined taxa: saprotrophic fungi (all other depicted fungal genera represent ectomycorrhizal fungi).



community composition (axis 1 = 41%). The variation explained by tree species was 13.8% ($P < 0.001$) in bacterial and 14.9% ($P < 0.001$) in fungal communities (Table S2). Furthermore, tree species (European beech or Norway spruce) had a stronger impact on soil bacterial and fungal community composition than soil depth, distance from tree trunk or season (Table S2). We identified specific indicator OTUs for soils surrounding beech or spruce stands (Table S5). Each bacterial indicator OTU showed an average relative abundance $< 1\%$, whereas few fungal indicator OTUs showed relative abundances $> 1\%$. Detailed information on relative abundances for all indicator OTUs is provided in Table S5.

For bacteria, 13 indicator OTUs were determined at the beech site and 10 indicator OTUs at the spruce site. The majority of bacterial OTUs representing indicators at the beech site were affiliated to Acidobacteria (mainly subgroup 2) (Table S5). Indicators at the spruce site comprised Chloroflexi, WD272 and several Acidobacteria subgroup 1 OTUs.

For both tree species, eight fungal OTUs were identified as potential indicators (Table S5). Under beech, a saprotrophic *Mortierella elongata* OTU and a *Trichoderma* OTU and ECM fungi OTUs (a *Russula cyanoxantha* OTU and a *Xerocomus chrysenteron* OTU) were identified as indicator OTUs. Indicators for spruce were three OTUs classified as saprotrophic fungi (*Exophiala* and two *Penicillium* OTUs). The two indicator ECM fungi under spruce were *Hygrophorus* and *Amphinema*.

Microbial community composition under both tree species was significantly affected by tree replicate, soil pH and OC (Table 2). Among the analyzed factors soil pH and tree species explained most of the variation in microbial community composition (Table S2).

Spatial and Seasonal Variability of Soil Microbial Community Composition

Bacterial community composition varied significantly with depth under spruce (Table 2). We found that relative abundance of OTUs of the dominant genus *Gaiella* was negatively correlated with OC concentration ($P < 0.001$) and higher at the 10- to 20-cm depth than the 0- to 10-cm depth. The relative abundance of the bacterial genus *Mucilaginibacter* also showed variations with soil depth. It was higher at the 0- to 10-cm depth vs. the 10- to 20-cm depth ($P < 0.001$). The fungal community composition showed no correlation with soil depth under both tree species (Table 2). However, the detected saprotrophic fungi were associated with the upper (0–10 cm depth) mineral soil layers, which were rich in OC (Figure 5). Additionally, the indicator species analysis identified mainly saprotrophic OTUs in the upper 10 cm of the studied soils (Table S5).

Spatial horizontal variation of overall bacterial community composition was significant in soil under beech ($P < 0.05$) and under spruce ($P < 0.001$) (Table 2). We found that relative abundance of the dominant bacterial genus *Nitrospira* was significantly higher at 3.5 m vs. 0.5 m distance from spruce trees. Furthermore, a Nitrospirales OTU was identified as an indicator for tree distances of 2.5 and 3.5 m (Table S5). Under beech trees, the relative abundance of *Pseudolabrys* differed significantly between 0.5 m and 3.5 m horizontal tree distance. Higher relative abundance was detected in soil located close to tree trunks. This effect was recorded with respect to both analyzed soil depths ($P < 0.05$). Overall fungal community composition differed significantly at different horizontal tree distances only in soil of the spruce stand ($P < 0.01$) (Table 2). However, fungal indicator species for certain combinations of

TABLE 2 | Multivariate analysis of variance based on weighted UniFrac distances with tree replicate, pH, OC, soil depth, season and distance as response variable.

	df	Beech stand				Spruce stand			
		Bacterial community		Fungal community		Bacterial community		Fungal community	
		MS	R ²	MS	R ²	MS	R ²	MS	R ²
Tree replicate	3	0.027	0.063**	1.637	0.233***	0.201	0.171***	2.227	0.323***
pH	1	0.221	0.171***	0.679	0.032**	0.385	0.109***	0.987	0.048***
OC	1	0.124	0.097***	0.843	0.04***	0.069	0.020	0.485	0.023**
Depth	1	0.016	0.013	0.277	0.013	0.100	0.028*	0.260	0.013
Season	1	0.078	0.06***	0.445	0.021	0.077	0.022	0.422	0.02*
Distance	1	0.026	0.021*	0.427	0.020	0.261	0.074***	0.672	0.032***
Residuals	55	0.014	0.576	0.245	0.640	0.037	0.575	0.204	0.541

Explanatory variables are given in rows in the order of entering the analysis. This table presents degrees of freedom (df), mean squares (MS), and R²-values. Significant results are indicated by *P < 0.05, **P < 0.01, ***P < 0.001.

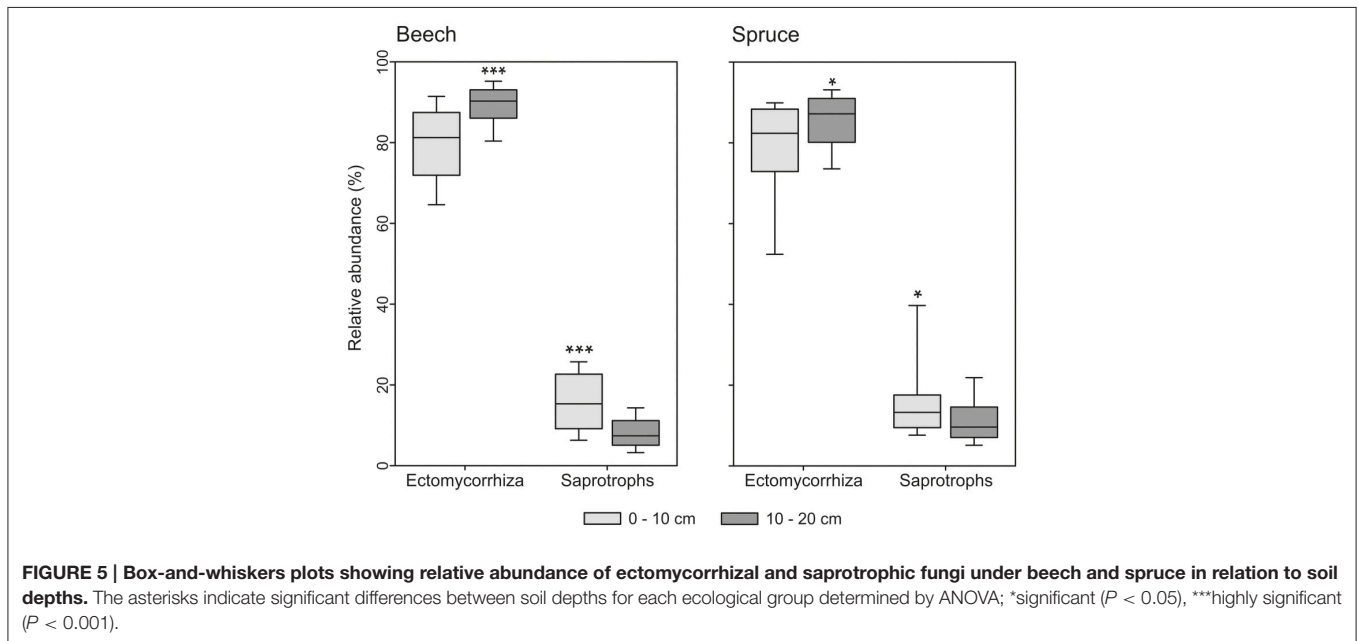


FIGURE 5 | Box-and-whiskers plots showing relative abundance of ectomycorrhizal and saprotrophic fungi under beech and spruce in relation to soil depths. The asterisks indicate significant differences between soil depths for each ecological group determined by ANOVA; *significant ($P < 0.05$), ***highly significant ($P < 0.001$).

tree distances were found in beech (Table S5) and spruce stands (Table S5).

A significant seasonal effect on bacterial community composition was detected in soil under beech ($P < 0.001$) (Table 2). Sequences corresponding to the Rhizobiales (*Bradyrhizobium* and *Rhodobium*) showed significantly higher relative abundance in autumn versus early summer ($P < 0.001$). Consistently, the analysis of indicator species identified an OTU affiliated to *Bradyrhizobium* in soil under beech in autumn (Table S5). A seasonal impact on fungal community composition was found in soil of the spruce stand ($P < 0.05$) (Table 2). Two fungal indicator species were identified in early summer in the spruce stand (Table S5). Fungal indicator species for both seasons (autumn and early summer) occurred under beech (Table S5).

DISCUSSION

Selective Association of Tree Species, Bacteria, and Fungi

Differences in distribution of microbial taxa were identified between soil under beech and spruce. This was expected, as even tree genotype within a species can have significant impacts on microbial communities (Schweitzer et al., 2008). A Chloroflexi OTU was identified as indicator for soil surrounding spruce. As several potential genes involved in phytochemical breakdown have been identified in Chloroflexi (Hug et al., 2013; Houghton et al., 2015), it is possible that this indicator microorganism plays a role in decomposition of spruce litter. Furthermore, the occurrence of several members of Acidobacteria was significantly affected by tree species. It can be assumed that acidobacterial

taxa contribute to decomposition in forest soils, as genomic and culture characteristics of subgroup 1 and 3 strains have been shown to utilize plant-derived biopolymers (Ward et al., 2009; García-Fraile et al., 2015). Shifts in occurrence of acidobacterial representatives between soil under European beech and Norway spruce might imply preferences for leaf or needle litter. A study on composition of bacterial communities under different deciduous and coniferous trees (e.g., *Picea* and Fagales species) in Czech forest stands also indicated litter preferences of Acidobacteria (Urbanová et al., 2015).

Forest vegetation (in particular dominant tree species) is important for distribution of mutualistic and saprotrophic fungi (Lauber et al., 2008; Goldmann et al., 2015). ECM fungi (e.g., *Russula*, *Inocybe*, *Piloderma*) establish mutualistic associations with plant roots (Smith and Read, 2008) and show preferences for particular tree species (Ishida et al., 2007; Thoms et al., 2010). In accordance with our study, Goldmann et al. (2015) and Miyamoto et al. (2015) reported a preference of *Tylospora* for coniferous trees. Some identified fungal indicators under beech (*Mortierella elongata*, *Trichoderma*, *Russula cyanoxantha*) are known to be widespread not just under a certain tree species (Wuczowski et al., 2003; Grebenc and Kraigher, 2007; Nagy et al., 2011). In contrast, the ECM fungus *Xerocomus chrysenteron* is known to have a preference for beech (Shi et al., 2002). Indicator species for spruce included three OTUs classified as saprotrophic fungi. *Exophiala* has already been described as a fungal genus decaying leaves in rainforests (Polishook et al., 1996) or existing as rhizospheric associates in temperate sites (Summerbell, 2005). Another two *Penicillium* OTUs were identified as saprotrophic indicators for spruce. Previous research (Johansson and Marklund, 1980) reported *Penicillium* to be antagonistic to *Fomes*, a well-known fungus infecting spruce trees (Schmidt, 2013). The indicative ECM fungi under spruce, *Hygrophorus* and *Amphinema*, were abundant and previously described for spruce ecosystems (Scattolin et al., 2008; Velmalá et al., 2013).

Under both tree species, microbial community composition was significantly affected by pH and OC concentration. Noteworthy, among the analyzed factors soil pH and tree species explained most of the variation in overall community composition of bacteria and fungi. Several previous studies have identified soil pH as a major driver of soil bacterial community composition across different regions and land use types (e.g., Lauber et al., 2009; Nacke et al., 2011). In accordance with our results, pH also explained a substantial fraction of variance in microbial community composition within other deciduous and coniferous forest soils (Lauber et al., 2009; Thoms et al., 2010; Goldmann et al., 2015). Furthermore, experiments including addition of substrates such as cellulose, lignin, and glucose to soil showed that the quantity of OC can have a significant impact on soil microbial community composition (Nakatsu et al., 2005; Goldfarb et al., 2011).

Relative Abundance of Saprotrophic Fungi Decreases with Soil Depth

Previous surveys based on DGGE analysis as well as Sanger sequencing and pyrosequencing of 16S rRNA genes have revealed differences in bacterial community composition between topsoil

and subsoil (Hansel et al., 2008; Eilers et al., 2012; Huang et al., 2013). This is a result of changes in soil characteristics such as organic C or N concentrations along soil profiles (Hansel et al., 2008; Will et al., 2010). Consistently, relative abundances of the bacterial genus *Gaiella*, which were higher in 10–20 cm depth than in 0–10 cm depth, were negatively correlated with organic C concentration. Different *Mucilaginibacter* representatives are capable of pectin, xylan, and laminarin degradation (Pankratov et al., 2007). *Mucilaginibacter* was more abundant in topsoils (0–10 cm). The genus has been previously associated with cellulose decomposition based on stable isotope probing (Štursová et al., 2012). Leaf and needle litter contains high amounts of the plant cell wall components xylan, pectin, and cellulose, and enters the upper mineral soil first, perhaps explaining the distribution of *Mucilaginibacter* OTUs.

Recently, McGuire et al. (2013) found discrete fungal communities in different soil horizons in boreal and tropical forest. This can be explained by changing carbon and nutrient contents in soil combined with fungal enzymatic decay abilities (McGuire et al., 2010; Prescott, 2010). Our results (Table 2) showed that fungal taxa in temperate forests do not underlay similar mechanisms as found previously. However, we identified different saprotrophic fungi showing preferences for the upper (0–10 cm depth) mineral soil layer, which was rich in OC. Influenced by the litter layer, the upper 10 cm show high habitat heterogeneity, competition amongst fungi for space, carbon and other soil nutrients (Kadowaki et al., 2014). ECM fungal taxa receive carbon through mycelium connected to plant roots (Smith and Read, 2008). In this study, ECM fungi were abundant irrespective of soil depth since these fungi are not C-limited and may colonize deeper soil layers (McGuire et al., 2013).

Bacteria Are Affected by horizontal Tree Distance under Beech and Spruce

Soil microbial community composition showed higher variability with respect to tree distance under spruce trees versus beech. It is known that spatial distribution of soil microbes can reflect the zone of influence and positioning of individual trees in forests (Saetre and Bååth, 2000; Ettema and Wardle, 2002). As stemflow was shown to significantly decrease soil pH, specifically close to beech trees (Koch and Matzner, 1993), we expected a clear change in microbial community composition next to beech trunks (0.5 m tree distance). However, we could neither detect a decrease in pH at 0.5 m distance to beech trunks, nor a strong change in microbial community composition next to the beech trees. Spatial horizontal variations in bacterial community composition under beech and spruce, recorded in this study, might have been partly evoked by changes in root activities with respect to varying tree distances. N demand of spruce trees in summer and autumn is mainly met by uptake of N compounds from soil and subsequent transport of reduced N from the roots to the shoot via the transpiration stream (Weber et al., 1998). Due to a negative relationship between fine root biomass and tree distance (steep decrease of fine root biomass at tree distances > 2 m) (Petritan et al., 2011), uptake of N compounds via roots might be more pronounced in soil located close to the analyzed coniferous tree trunks. This potentially explains the spatial horizontal variations

in occurrence of nitrifying bacteria belonging to Nitrospirales under spruce.

Under beech, relative abundance of *Pseudolabrys* was significantly affected by horizontal tree distance. Only one *Pseudolabrys* species, isolated from Taiwanese soil, has been described (Kämpfer et al., 2006). In our study, more than one OTU determined at a genetic distance of 3% was affiliated to *Pseudolabrys*. The taxon *Pseudolabrys*, representing one of the most abundant genera detected in this study, belongs to the Rhizobiales, which are known to interact with plants (Erlacher et al., 2015). Changes in root densities or activities may be a major reason for high relative abundance of *Pseudolabrys* in soil located close to beech trunks.

Branco et al. (2013) found that an increase in soil pH with pine tree distance was related to changing occurrence of fungal species. Variation in pH at different tree distances (Table S4) also account for changes in fungal community composition under the conifer trees analyzed in our study ($P < 0.05$) (Table 2).

More Seasonal Soil Community Variation in Beech than in Spruce Forests

Soil bacterial community composition under beech was strongly affected by season ($P < 0.001$). Recently, López-Mondéjar et al. (2015) reported that bacterial communities undergo seasonal changes in mineral soil of a *Quercus petraea* (Matt.) Liebl forest. They assume that seasonal differences in the activity of tree roots are a major driver of soil bacterial community composition in deciduous forest. Here, we found that different members of the Rhizobiales were more abundant under beech in autumn than in early summer. As Rhizobiales are known to interact with plants, seasonal root impacts might affect their abundance in temperate deciduous forest. Understory vegetation varies between European beech and Norway spruce age class forests in the study region (Boch et al., 2013). It is possible that the Rhizobiales community is affected by seasonal changes in understory vegetation. Furthermore, seasonal shifts in soil moisture and temperature may also affect bacterial community composition in the analyzed soil (Kaiser et al., 2010; Shay et al., 2015).

Seasonal impacts on fungi were reported previously (e.g., Stevenson et al., 2014; Moll et al., 2015). In this study, soil fungal community composition was affected by season under spruce ($P < 0.05$) but not as expected under beech. Recently, Voříšková et al. (2014) also detected no significant seasonal effect on fungal community composition in soil of a deciduous forest (oak forest near Prague, Czech Republic). Nevertheless, in the litter horizon, which was not analyzed in our study, seasonal changes in fungal community composition were identified by Voříšková et al. (2014). These changes are associated with nutrient input from fresh litter, which occurs in temperate deciduous forests each autumn (Voříšková et al., 2014). In accordance with our study, Lin et al. (2016) reported seasonal shifts of fungi in coniferous forests. The air and soil temperatures at both forest stands were higher in early summer, whereas the soil water content was increased in autumn (Table S6). Hence, comparable weather conditions

would suggest similar fungal reactions toward changing season at the beech and spruce stand. However, a relatively thick needle litter layer (~8 cm) was removed before soil sampling under spruce. Breakdown of needles, which are highly recalcitrant to biological degradation, is mainly performed by fungi. It is possible that the distinct fungi colonizing needles (Korkama-Rajala et al., 2008) and consequently soil fungal communities under coniferous trees are susceptible to climatic changes in autumn. In addition, unmeasured factors might account for the shifts of fungal communities under spruce. Future studies can evaluate if these findings are artificial or ecologically reasonable.

CONCLUSION

In accordance with our first hypothesis, beech and spruce trees strongly shaped the community composition of soil bacteria and fungi in temperate forests. Tree species-specific preferences with respect to bacterial and fungal microorganisms, such as a Chloroflexi representative, members of Acidobacteria subgroup 2 or *Hygrophorus* and *Clavulina*, were identified. Trees also have manifold impacts on the seasonal and spatial distribution of soil microorganisms. Indicator species analyses showed a vertical variation with a higher importance of saprotrophic taxa in the upper soil layer (0–10 cm) compared to the soil at a depth of 10–20 cm, supporting our second hypothesis. In line with our third hypothesis, bacterial community composition was strongly affected by tree distance, which might be due to higher fine root biomass near spruce trunks. Furthermore, bacterial community composition showed stronger seasonal variation under deciduous trees versus evergreen trees. This pattern was not found when analyzing fungal community composition, which is in contrast to our forth hypothesis. Noteworthy, soil fungal communities under spruce seem to be susceptible to seasonal changes. Overall, our results indicate that trees influence the spatial variation of bacteria and fungi, but their diverse patterns in stem flow, measured by pH change, seem to have a minor impact. Furthermore, the study indicates that soil pH and tree species (European beech or Norway spruce) have a stronger impact on soil bacterial and fungal community composition than soil depth, season or distance from tree trunk.

Additional studies considering root architecture and exudation patterns as well as the influence of tree canopy on the spatial distribution of leaf litter fall are necessary to further elucidate interactions between trees and soil microbes. Besides studies allowing analysis of the proportional importance of factors such as tree species, tree distance, or season, and their mechanisms for interaction, experimental designs focusing on effects of single factors are required to gain more comprehensive understanding on microbial community variation in forest soil. Furthermore, more direct proof is needed to ascertain functional roles of microbes such as Acidobacteria in soil surrounding beech and spruce. For instance, stable isotope probing could be used to identify bacteria or fungi involved in litter degradation.

AUTHOR CONTRIBUTIONS

MS, FB, RD, and TW designed the study; HN, KG, IS, BP, KK, and GC carried out field and laboratory work; HN, KG, IS, and KK prepared and analyzed the data; all authors interpreted the results and wrote the paper.

ACKNOWLEDGMENTS

We thank the managers of the three Exploratories, Kirsten Reichel-Jung, Swen Renner, Katrin Hartwich, Sonja Gockel, Kerstin Wiesner, and Martin Gorke for their work in maintaining the plot and project infrastructure; Christiane Fischer and Simone Pfeiffer for giving support through the central office, Michael Owonibi for managing the central data base, and Markus Fischer, Eduard Linsenmair, Dominik Hessenmöller, Jens Nieschulze, Daniel Prati, Ernst-Detlef Schulze, Wolfgang W. Weisser, and the late Elisabeth Kalko for their role in setting up the Biodiversity Exploratories project. The work was funded by the DFG Priority Program 1374 “Infrastructure-Biodiversity-Exploratories” (DA 374/6-1, Core Project 10—SCHR 1181/2-2 and Core Project 11—BU 941/22-1). Field work permits were issued by the responsible state environmental office of Thüringen (according to § 72 BbgNatSchG). This work was further supported by Helmholtz Impulse and Networking Fund through Helmholtz Interdisciplinary Graduate School for Environmental Research (HIGRADE). We kindly thank Beatrix Schnabel for her help while sampling, lab work and together with Melanie Günther and Sigrid Härtling for 454 sequencing. Furthermore, we thank Sandra Klemmer, Theresa Klötzing, Jessica Schäfer, Frederic Van Broeck, Maximilian Frei, and Steffen Both for their field and/or lab assistance and Carla Porges for the drawings

REFERENCES

- Augusto, L., Ranger, J., Binkley, D., and Rothe, A. (2002). Impact of several common tree species of European temperate forests on soil fertility. *Ann. For. Sci.* 59, 233–253. doi: 10.1051/forest:2002020
- Ayres, E., Steltzer, H., Berg, S., Wallenstein, M. D., Simmons, B. L., and Wall, D. H. (2009). Tree species traits influence soil physical, chemical, and biological properties in high elevation forests. *PLoS ONE* 4:e5964. doi: 10.1371/journal.pone.0005964
- Baldrian, P., Kolařík, M., Stursová, M., Kopecký, J., Valášková, V., Větrovský, T., et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 6, 248–258. doi: 10.1038/ismej.2011.95
- Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2015). GenBank. *Nucleic Acids Res.* 43, D30–D35. doi: 10.1093/nar/gku1216
- Berger, T. W., and Berger, P. (2012). Greater accumulation of litter in spruce (*Picea abies*) compared to beech (*Fagus sylvatica*) stands is not a consequence of the inherent recalcitrance of needles. *Plant Soil* 358, 349–369. doi: 10.1007/s11104-012-1165-z
- Boch, S., Prati, D., Müller, J., Socher, S., Baumbach, H., Buscot, F., et al. (2013). High plant species richness indicates management-related disturbances rather than the conservation status of forests. *Basic Appl. Ecol.* 14, 496–505. doi: 10.1016/j.baee.2013.06.001
- Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G. W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9, 425–426. doi:10.1038/nmeth.1990
- Branco, S., Bruns, T. D., and Singleton, I. (2013). Fungi at a small scale: spatial zonation of fungal assemblages around single trees. *PLoS ONE* 8:e78295. doi: 10.1371/journal.pone.0078295
- Broszat, M., Nacke, H., Blasi, R., Siebe, C., Huebner, J., Daniel, R., et al. (2014). Wastewater irrigation increases the abundance of potentially harmful gammaproteobacteria in soils in Mezquital Valley, Mexico. *Appl. Environ. Microbiol.* 80, 5282–5291. doi: 10.1128/AEM.01295-14
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Cesarz, S., Fender, A. C., Beyer, F., Valtanen, K., Pfeiffer, B., Gansert, D., et al. (2013). Roots from beech (*Fagus sylvatica* L.) and ash (*Fraxinus excelsior* L.) differentially affect soil microorganisms and carbon dynamics. *Soil Biol. Biochem.* 61, 23–32. doi:10.1016/j.soilbio.2013.02.003
- Coleman, D. C., and Crossley, D. A. (1996). *Fundamentals of Soil Ecology*. New York, NY: Academic Press.
- Cong, J., Yang, Y., Liu, X., Lu, H., Liu, X., Zhou, J., et al. (2015). Analyses of soil microbial community compositions and functional genes reveal potential consequences of natural forest succession. *Sci. Rep.* 5:10007. doi: 10.1038/srep10007
- Crowther, T. W., Glick, H. B., Covey, K. R., Bettigole, C., Maynard, D. S., Thomas, S. M., et al. (2015). Mapping tree density at a global scale. *Nature* 525, 201–205. doi: 10.1038/nature14967
- De Cáceres, M., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

of our sampling design. Additionally, we acknowledge support by the DFG and the Open Access Publication Funds of the Göttingen University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.02067/full#supplementary-material>

Table S1 | Information on selected trees including tree position, tree age, tree trunk circumference, and sampling direction away from the tree trunk in May and November 2012.

Table S2 | Multivariate analysis of variance based on weighted UniFrac distances of whole bacterial and fungal community composition.

Table S3 | Multivariate analysis of variance based on weighted UniFrac distances of bacterial and fungal community composition under beech or spruce.

Table S4 | Analysis of covariance to test the impact of tree replicate, depth, season and distance from the tree trunk on soil chemical parameters under trees.

Table S5 | Results of indicator species analysis showing potential indicative OTUs for tree species, sampling depths, sampling distances and sampling season under trees.

Table S6 | Air temperature, soil temperature, and soil water content data.

Figure S1 | Rarefaction curves indicating the observed numbers of OTUs at a genetic distance of 3%. Samples derived from soil surrounding beech and samples collected under spruce are represented by brown and green color, respectively.

Figure S2 | Line plots showing soil bacterial and fungal diversity as assessed by Shannon index at 3% genetic distance under beech in (A) early summer and (C) autumn, and under spruce in (B) early summer and (D) autumn.

- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- Eilers, K. G., Debenport, S., Anderson, S., and Fierer, N. (2012). Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biol. Biochem.* 50, 58–65. doi: 10.1016/j.soilbio.2012.03.011
- Erlacher, A., Cernava, T., Cardinale, M., Soh, J., Sensen, C. W., Grube, M., et al. (2015). *Rhizobiales* as functional and endosymbiotic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Front. Microbiol.* 6:53. doi: 10.3389/fmicb.2015.00053
- Ettema, C. H., and Wardle, D. A. (2002). Spatial soil ecology. *Trends Ecol. Evol.* 17, 177–183. doi: 10.1016/S0169-5347(02)02496-5
- Fender, A. C., Gansert, D., Jungkunst, H. F., Fiedler, S., Beyer, F., Schützenmeister, K., et al. (2013). Root-induced tree species effects on the source/sink strength for greenhouse gases (CH₄, N₂O and CO₂) of a temperate deciduous forest soil. *Soil Biol. Biochem.* 57, 587–597. doi: 10.1016/j.soilbio.2012.08.004
- Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., et al. (2010). Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. *Basic Appl. Ecol.* 11, 473–485. doi: 10.1016/j.baee.2010.07.009
- García-Fraile, P., Benada, O., Cajthaml, T., Baldrian, P., and Lladó, S. (2015). *Terracidiphilus gabretensis* gen. nov., sp. nov., an abundant and active forest soil Acidobacterium important in organic matter transformation. *Appl. Environ. Microbiol.* 82, 560–569. doi: 10.1128/AEM.03353-15
- Gardes, M., and Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Mol. Ecol.* 2, 113–118. doi: 10.1111/j.1365-294X.1993.tb00005.x
- Gefler, A., Schneider, S., Weber, P., Hanemann, U., and Rennenberg, H. (1998). Soluble N compounds in trees exposed to high loads of N: a comparison between the roots of Norway spruce (*Picea abies*) and beech (*Fagus sylvatica*) trees grown under field conditions. *New Phytol.* 138, 385–399. doi: 10.1046/j.1469-8137.1998.00134.x
- Goldfarb, K. C., Karaoz, U., Hanson, C. A., Santee, C. A., Bradford, M. A., Treseder, K. K., et al. (2011). Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front. Microbiol.* 2:94. doi: 10.3389/fmicb.2011.00094
- Goldmann, K., Schöning, I., Bucot, F., and Wubet, T. (2015). Forest management type influences diversity and community composition of soil fungi across temperate forest ecosystems. *Front. Microbiol.* 6:1300. doi: 10.3389/fmicb.2015.01300
- Grebenc, T., and Kraigher, H. (2007). Types of ectomycorrhiza of mature beech and spruce at ozone-fumigated and control forest plots. *Environ. Monit. Assess.* 128, 47–59. doi: 10.1007/s10661-006-9414-3
- Hammer, Ø., Harper, D. A. T., and Ryan, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4, 1–9. Available online at: http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- Hanewinkel, M., Cullmann, D. A., Schelhaas, M. J., Nabuurs, G. J., and Zimmermann, N. E. (2013). Climate change may cause severe loss in the economic value of European forest land. *Nat. Clim. Change* 3, 203–207. doi: 10.1038/nclimate1687
- Hansel, C. M., Fendorf, S., Jardine, P. M., and Francis, C. A. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl. Environ. Microbiol.* 74, 1620–1633. doi: 10.1128/AEM.01787-07
- Houghton, K. M., Morgan, X. C., Lagutin, K., MacKenzie, A. D., Vyssotskii, M., Mitchell, K. A., et al. (2015). *Thermorudis pharmacophila* sp. nov., a novel member of the class Thermomicrobia isolated from geothermal soil, and emended descriptions of *Thermomicrobium roseum*, *Thermomicrobium carboxidum*, *Thermorudis peleae* and *Sphaerobacter thermophilus*. *Int. J. Syst. Evol. Microbiol.* 65, 4479–4487. doi: 10.1099/ijsem.000598
- Huang, J., Sheng, X., He, L., Huang, Z., Wang, Q., and Zhang, Z. (2013). Characterization of depth-related changes in bacterial community compositions and functions of a paddy soil profile. *FEMS Microbiol. Lett.* 347, 33–42. doi: 10.1111/1574-6968.12218
- Hug, L. A., Castelle, C. J., Wrighton, K. C., Thomas, B. C., Sharon, I., Frischkorn, K. R., et al. (2013). Community genomic analyses constrain the distribution of metabolic traits across the *Chloroflexi phylum* and indicate roles in sediment carbon cycling. *Microbiome* 1:22. doi: 10.1186/2049-2618-1-22
- Ishida, T. A., Nara, K., and Hogetsu, T. (2007). Host effects on ectomycorrhizal fungal communities: insight from eight host species in mixed conifer-broadleaf forests. *New Phytol.* 174, 430–440. doi: 10.1111/j.1469-8137.2007.02016.x
- Jiang, Y., Chen, C., Xu, Z., and Liu, Y. (2012). Effects of single and mixed species forest ecosystems on diversity and function of soil microbial community in subtropical China. *J. Soils Sediments* 12, 228–240. doi: 10.1007/s11368-011-0442-4
- Johansson, M., and Marklund, E. (1980). Antagonists of *Fomes annosus* in the rhizosphere of grey alder (*Alnus incana*) and Norway spruce (*Picea abies*). *Eur. J. For. Pathol.* 10, 385–395. doi: 10.1111/j.1439-0329.1980.tb00056.x
- Johnson, M. S., and Lehmann, J. (2006). Double-funneling of trees: stemflow and root-induced preferential flow. *Ecoscience* 13, 324–333. doi: 10.2980/i1195-6860-13-3-324.1
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201
- Kadowaki, K., Sato, H., Yamamoto, S., Tanabe, A. S., Hidaka, A., and Toju, H. (2014). Detection of the horizontal spatial structure of soil fungal communities in a natural forest. *Popul. Ecol.* 56, 301–310. doi: 10.1007/s10144-013-0424-z
- Kaiser, C., Koranda, M., Kitzler, B., Fuchsluger, L., Schneckler, J., Schweiger, P., et al. (2010). Belowground carbon allocation by trees drives seasonal patterns of extracellular enzyme activities by altering microbial community composition in a beech forest soil. *New Phytol.* 187, 843–858. doi: 10.1111/j.1469-8137.2010.03321.x
- Kämpfer, P., Young, C. C., Arun, A. B., Shen, F. T., Jäckel, U., Rosselló-Mora, R., et al. (2006). *Pseudolabrys taiwanensis* gen. nov., sp. nov., an alphaproteobacterium isolated from soil. *Int. Syst. Evol. Microbiol.* 56, 2469–2472. doi: 10.1099/ijss.0.64124-0
- Koch, A. S., and Matzner, E. (1993). Heterogeneity of soil and soil solution chemistry under Norway Spruce (*Picea abies* Karst.) and European Beech (*Fagus sylvatica* L.) as influenced by distance from the stem basis. *Plant Soil* 151, 227–237. doi: 10.1007/BF00016288
- Köljal, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. doi: 10.1111/mec.12481
- Korkama-Rajala, T., Müller, M. M., and Pennanen, T. (2008). Decomposition and fungi of needle litter from slow- and fast-growing Norway spruce (*Picea abies*) clones. *Microb. Ecol.* 56, 76–89. doi: 10.1007/s00248-007-9326-y
- Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75, 5111–5120. doi: 10.1128/AEM.00335-09
- Lauber, C. L., Strickland, M. S., Bradford, M. A., and Fierer, N. (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* 40, 2407–2415. doi: 10.1016/j.soilbio.2008.05.021
- Lejon, D. P., Chaussod, R., Ranger, J., and Ranjard, L. (2005). Microbial community structure and density under different tree species in an acid forest soil (Morvan, France). *Microb. Ecol.* 50, 614–625. doi: 10.1007/s00248-005-5130-8
- Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Lin, W. R., Wang, P. H., Chen, W. C., Lai, C. M., and Winder, R. S. (2016). Responses of soil fungal populations and communities to the thinning of *Cryptomeria japonica* Forests. *Microbes Environ.* 31, 19–26. doi: 10.1264/jisme2.ME15127
- Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D., and Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* 35:e120. doi: 10.1093/nar/gkm541
- Štursová, M., Žifčáková, L., Leigh, M. B., Burgess, R., and Baldrian, P. (2012). Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol. Ecol.* 80, 735–746. doi: 10.1111/j.1574-6941.2012.01343.x

- López-Mondéjar, R., Voříšková, J., Větrovský, T., and Baldrian, P. (2015). The bacterial community inhabiting temperate deciduous forests is vertically stratified and undergoes seasonal dynamics. *Soil Biol. Biochem.* 87, 43–50. doi: 10.1016/j.soilbio.2015.04.008
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/embnet.17.1.200
- McGuire, K. L., Allison, S. D., Fierer, N., and Treseder, K. K. (2013). Ectomycorrhizal-dominated boreal and tropical forests have distinct fungal communities, but analogous spatial patterns across soil horizons. *PLoS ONE* 8:e68278. doi: 10.1371/journal.pone.0068278
- McGuire, K. L., Bent, E., Borneman, J., Majumder, A., Allison, S. D., and Treseder, K. K. (2010). Functional diversity in resource use by fungi. *Ecology* 91, 2324–2332. doi: 10.1890/09-0654.1
- Miyamoto, Y., Sakai, A., Hattori, M., and Nara, K. (2015). Strong effect of climate on ectomycorrhizal fungal composition: evidence from range overlap between two mountains. *ISME J.* 9, 1870–1879. doi: 10.1038/ismej.2015.8
- Moll, J., Goldmann, K., Kramer, S., Hempel, S., Kandel, E., Marhan, S., et al. (2015). Resource type and availability regulate fungal communities along arable soil profiles. *Microb. Ecol.* 70, 390–399. doi: 10.1007/s00248-015-0569-8
- Morales, S. E., Cosart, T. F., Johnson, J. V., and Holben, W. E. (2009). Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. *Appl. Environ. Microbiol.* 75, 668–675. doi: 10.1128/AEM.01757-08
- Nacke, H., Thürmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., et al. (2011). Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS ONE* 6:e17000. doi: 10.1371/journal.pone.0017000
- Nagy, L. G., Petkovits, T., Kovacs, G. M., Voigt, K., Vagvolgyi, C., and Papp, T. (2011). Where is the unseen fungal diversity hidden? A study of Mortierella reveals a large contribution of reference collections to the identification of fungal environmental sequences. *New Phytol.* 191, 789–794. doi: 10.1111/j.1469-8137.2011.03707.x
- Nakatsu, C. H., Carmosini, N., Baldwin, B., Beasley, F., Kourtev, P., and Konopka, A. (2005). Soil microbial community responses to additions of organic carbon substrates and heavy metals (Pb and Cr). *Appl. Environ. Microbiol.* 71, 7679–7689. doi: 10.1128/AEM.71.12.7679-7689.2005
- Nykvist, N. (1963). Leaching and decomposition of water-soluble organic substances from different types of leaf and needle litter. *Stud. For. Suec.* 3, 1–31.
- Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., et al. (2016). *vegan: Community Ecology Package*. Oulu: University of Oulu. Available online at: <http://CRAN.R-project.org/package=vegan>
- Pankratov, T. A., Tindall, B. J., Liesack, W., and Dedysh, S. N. (2007). *Mucilaginibacter paludis* gen. nov., sp. nov. and *Mucilaginibacter gracilis* sp. nov., pectin-, xylan- and laminarin-degrading members of the family Sphingobacteriaceae from acidic Sphagnum peat bog. *Int. J. Syst. Evol. Microbiol.* 57, 2349–2354. doi: 10.1099/ijs.0.65100-0
- Petritan, I. C., von Lupke, B., and Petritan, A. M. (2011). Fine roots of overstory Norway spruce (*Picea abies*): distribution and influence on growth of underplanted beech (*Fagus sylvatica*) and Douglas-fir (*Pseudotsuga menziesii*) saplings. *For. Syst.* 20, 407–419. doi: 10.5424/fs/20112003/11136
- Pfeiffer, B., Fender, A. C., Lasota, S., Hertel, D., Jungkunst, H. F., and Daniel, R. (2013). Leaf litter is the main driver for changes in bacterial community structures in the rhizosphere of ash and beech. *Appl. Soil Ecol.* 72, 150–160. doi: 10.1016/j.apsoil.2013.06.008
- Polishook, J. D., Bills, G. F., and Lodge, D. J. (1996). Microfungi from decaying leaves of two rain forest trees in Puerto Rico. *J. Ind. Microbiol. Biotechnol.* 17, 284–294. doi: 10.1007/BF01574703
- Prescott, C. E. (2010). Litter decomposition: what controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry* 101, 133–149. doi: 10.1007/s10533-010-9439-0
- Priha, O., and Smolander, A. (1997). Microbial biomass and activity in soil and litter under *Pinus sylvestris*, *Picea abies* and *Betula pendula* at originally similar afforestation sites. *Biol. Fert. Soils* 24, 45–51. doi: 10.1007/BF01420219
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi: 10.1093/nar/gkm864
- Purahong, W., Hoppe, B., Kahl, T., Schlotter, M., Schulze, E. D., Bauhus, J., et al. (2014). Changes within a single land-use category alter microbial diversity and community structure: molecular evidence from wood-inhabiting fungi in forest ecosystems. *J. Environ. Manage.* 139, 109–119. doi: 10.1016/j.jenvman.2014.02.031
- Raz-Yaseef, N., Rotenberg, E., and Yakir, D. (2010). Effects of spatial variations in soil evaporation caused by tree shading on water flux partitioning in a semi-arid pine forest. *Agric. For. Meteorol.* 150, 454–462. doi: 10.1016/j.agrformet.2010.01.010
- R Development Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing.
- Rosselló-Mora, R., and Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67. doi: 10.1111/j.1574-6976.2001.tb00571.x
- Saetre, P., and Bååth, E. (2000). Spatial variation and patterns of soil microbial community structure in a mixed spruce-birch stand. *Soil Biol. Biochem.* 32, 909–917. doi: 10.1016/S0038-0717(99)00215-1
- Scattolin, L., Montecchio, L., Mosca, E., and Agerer, R. (2008). Vertical distribution of the ectomycorrhizal community in the top soil of Norway spruce stands. *Eur. J. For. Res.* 127, 347–357. doi: 10.1007/s10342-008-0209-7
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Schmidt, O. (2013). *Holz-und Baumpilze: Biologie, Schäden, Schutz, Nutzen*. Heidelberg: Springer-Verlag.
- Schweitzer, J. A., Bailey, J. K., Fischer, D. G., LeRoy, C. J., Lonsdorf, E. V., Whitham, T. G., et al. (2008). Plant-soil-microorganism interactions: heritable relationship between plant genotype and associated soil microorganisms. *Ecology* 89, 773–781. doi: 10.1890/07-0337.1
- Shay, P. E., Winder, R. S., and Trofymow, J. A. (2015). Nutrient-cycling microbes in coastal Douglas-fir forests: regional-scale correlation between communities, *in situ* climate, and other factors. *Front. Microbiol.* 6:1097. doi: 10.3389/fmicb.2015.01097
- Shi, L., Guttenberger, M., Kottke, I., and Hampp, R. (2002). The effect of drought on mycorrhizas of beech (*Fagus sylvatica* L.): changes in community structure, and the content of carbohydrates and nitrogen storage bodies of the fungi. *Mycorrhiza* 12, 303–311. doi: 10.1007/s00572-002-0197-2
- Smith, S. E., and Read, D. J. (2008). *Mycorrhizal Symbiosis*. London: Academic press, Elsevier.
- Stevenson, B. A., Hunter, D. W. F., and Rhodes, P. L. (2014). Temporal and seasonal change in microbial community structure of an undisturbed, disturbed, and carbon-amended pasture soil. *Soil Biol. Biochem.* 75, 175–185. doi: 10.1016/j.soilbio.2014.04.010
- Summerbell, R. C. (2005). Root endophyte and mycorrhizosphere fungi of black spruce, *Picea mariana*, in a boreal forest habitat: influence of site factors on fungal distributions. *Stud. Mycol.* 53, 121–145. doi: 10.3114/sim.53.1.121
- Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., et al. (2016). Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *ISME J.* 10, 346–362. doi: 10.1038/ismej.2015.116
- Thoms, C., Gättinger, A., Jacob, M., Thomas, F. M., and Gleixner, G. (2010). Direct and indirect effects of tree diversity drive soil microbial diversity in temperate deciduous forest. *Soil Biol. Biochem.* 42, 1558–1565. doi: 10.1016/j.soilbio.2010.05.030
- Thoms, C., and Gleixner, G. (2013). Seasonal differences in tree species' influence on soil microbial communities. *Soil Biol. Biochem.* 66, 239–248. doi: 10.1016/j.soilbio.2013.05.018
- Trevors, J. T. (2010). One gram of soil: a microbial biochemical gene library. *Antonie van Leeuwenhoek* 97, 99–106. doi: 10.1007/s10482-009-9397-5
- Urbanová, M., Šnajdr, J., and Baldrian, P. (2015). Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. *Soil Biol. Biochem.* 84, 53–64. doi: 10.1016/j.soilbio.2015.02.011

- Uroz, S., Oger, P., Tisserand, E., Cébron, A., Turpault, M. P., Buée, M., et al. (2016). Specific impacts of beech and Norway spruce on the structure and diversity of the rhizosphere and soil microbial communities. *Sci. Rep.* 6:27756. doi: 10.1038/srep27756
- van der Heijden, M. G., Bardgett, R. D., and van Straalen, N. M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* 11, 296–310. doi: 10.1111/j.1461-0248.2007.01139.x
- Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., and Knight, R. (2013). Emperor: a tool for visualizing high-throughput microbial community data. *Gigascience* 2:16. doi: 10.1186/2047-217X-2-16
- Velmalá, S. M., Rajala, T., Haapanen, M., Taylor, A. F., and Pennanen, T. (2013). Genetic host-tree effects on the ectomycorrhizal community and root characteristics of Norway spruce. *Mycorrhiza* 23, 21–33. doi: 10.1007/s00572-012-0446-y
- Voríšková, J., Brabcová, V., Cajthaml, T., and Baldrian, P. (2014). Seasonal dynamics of fungal communities in a temperate oak forest soil. *New Phytol.* 201, 269–278. doi: 10.1111/nph.12481
- Wäldchen, J., Schöning, I., Mund, M., Schrumpf, M., Bock, S., Herold, N., et al. (2012). Estimation of clay content from easily measurable water content of air-dried soil. *J. Plant Nutr. Soil Sci.* 175, 367–376. doi: 10.1002/jpln.201100066
- Wäldchen, J., Schulze, E. D., Mund, M., and Winkler, B. (2011). Der Einfluss politischer, rechtlicher und wirtschaftlicher Rahmenbedingungen des 19. Jahrhunderts auf die Bewirtschaftung der Wälder im Hainich-Dün-Gebiet (Nordthüringen). *Forstarchiv* 82, 35–47. doi: 10.2376/0300-4112-82-35
- Wang, Y., and Qian, P. Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4:e7401. doi: 10.1371/journal.pone.0007401
- Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., et al. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* 75, 2046–2056. doi: 10.1128/AEM.02294-08
- Weber, P., Stoermer, H., Geßler, A., Schneider, S., von Sengbusch, D., Hanemann, U., et al. (1998). Metabolic responses of Norway spruce (*Picea abies*) trees to long-term forest management practices and acute (NH₄)₂SO₄ fertilization: transport of soluble non-protein nitrogen compounds in xylem and phloem. *New Phytol.* 140, 461–475. doi: 10.1111/j.1469-8137.1998.00285.x
- White, T., Brans, T., Lee, S., and Taylor, J. (1990). “Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics,” in *PCR Protocols: A Guide to Methods and Applications*, eds M. A. Innis, D. H. Gelfand and J. J. Sninsky (San Diego, CA: Academic Press), 315.
- Will, C., Thürmer, A., Wollherr, A., Nacke, H., Herold, N., Schrumpf, M., et al. (2010). Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Appl. Environ. Microbiol.* 76, 6751–6759. doi: 10.1128/AEM.01063-10.
- Wubet, T., Christ, S., Schöning, I., Boch, S., Gawlich, M., Schnabel, B., et al. (2012). Differences in soil fungal communities between European beech (*Fagus sylvatica* L.) dominated forests are related to soil and understory vegetation. *PLoS ONE* 7:e47500. doi: 10.1371/journal.pone.0047500
- Wuczkowski, M., Druzhinina, I., Gherbawy, Y., Klug, B., Prillinger, H., and Kubicek, C. P. (2003). Species pattern and genetic diversity of Trichoderma in a mid-European, primeval floodplain-forest. *Microbiol. Res.* 158, 125–133. doi: 10.1078/0944-5013-00193
- Yarwood, S. A., Bottomley, P. J., and Myrold, D. D. (2010). Soil microbial communities associated with Douglas-fir and red alder stands at high- and low-productivity forest sites in Oregon, USA. *Microb. Ecol.* 60, 606–617. doi: 10.1007/s00248-010-9675-9
- Zechmeister-Boltenstern, S., Michel, K., and Pfeffer, M. (2011). Soil microbial community structure in European forests in relation to forest type and atmospheric nitrogen deposition. *Plant Soil* 343, 37–50. doi: 10.1007/s11104-010-0528-6
- Zeng, Q., Dong, Y., and An, S. (2016). Bacterial community responses to soils along a latitudinal and vegetation gradient on the Loess Plateau, China. *PLoS ONE* 11:e0152894. doi: 10.1371/journal.pone.0152894

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nacke, Goldmann, Schöning, Pfeiffer, Kaiser, Castillo-Villamizar, Schrumpf, Buscot, Daniel and Wubet. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplemental Information for Chapter 8.2.1

Contents

Table S1. Information on selected trees including tree position, tree age, tree trunk circumference, and sampling direction away from the tree trunk in May and November 2012.

Table S2. Multivariate analysis of variance based on weighted UniFrac distances of whole bacterial and fungal community composition.

Table S3. Multivariate analysis of variance based on weighted UniFrac distances of bacterial and fungal community composition under beech or spruce.

Table S4. Analysis of covariance to test the impact of tree replicate, depth, season and distance from the tree trunk on soil chemical parameters under trees.

Table S5. Results of indicator species analysis showing potential indicative OTUs for tree species, sampling depths, sampling distances and sampling season under trees.

Table S6. Air temperature, soil temperature, and soil water content data.

Figure S1. Rarefaction curves indicating the observed numbers of OTUs at a genetic distance of 3%. Samples derived from soil surrounding beech and samples collected under spruce are represented by brown and green color, respectively.

Figure S2. Line plots showing soil bacterial and fungal diversity as assessed by Shannon index at 3% genetic distance under beech in (A) early summer and (C) autumn, and under spruce in (B) early summer and (D) autumn.

Table S1. Information on selected trees including tree position, tree age, tree trunk circumference, and sampling direction away from the tree trunk in May and November 2012.

Tree species	Replicate	Easting [m]	Northing [m]	Direction		Tree age [a]	Trunk circumference (h=1.50m) [m]
				1st sampling Early summer [°]	2nd sampling Autumn [°]		
<i>Fagus sylvatica</i>	1	3586543±5	5682178±5	135	118	55-65	1.56
<i>Fagus sylvatica</i>	2	3586530±5	5682187±5	182	165	55-65	1.2
<i>Fagus sylvatica</i>	3	3585534±5	5682178±5	293	285	55-65	1.34
<i>Fagus sylvatica</i>	4	3586530±5	5682162±5	35	20	55-65	1.4
<i>Picea abies</i>	1	3591581±7	5682660±7	195	213	50-60	1.38
<i>Picea abies</i>	2	3591567±7	5682668±7	350	5	50-60	1.09
<i>Picea abies</i>	3	3591595±7	5682693±7	5	0	50-60	1.22
<i>Picea abies</i>	4	3591594±7	5682652±7	210	195	50-60	1.26

Table S2. Multivariate analysis of variance based on weighted UniFrac distances of whole bacterial and fungal community composition. (a) Stepwise model testing for multivariate analysis of variance based on weighted UniFrac distances of whole bacterial and fungal community composition (128 samples) with seven considered response factors in the order of entering the analysis. Increasing R^2 -values, and adjusted R^2 -values represent chosen model suitability. (b) Analysis of final model with seven response variables. R^2 and adjusted R^2 for each variable within the model. Explanatory variables are given in rows in the order of entering the analysis. Significant results are indicated by * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Abbreviation: Adj., Adjusted.

(a)

No. of factors	Factor	Bacterial community		Fungal community	
		R^2	Adj. R^2	R^2	Adj. R^2
1	Replicate	0.064	0.056	0.107	0.100
2	pH	0.206	0.193	0.166	0.153
3	OC	0.247	0.229	0.185	0.165
4	Depth	0.265	0.241	0.200	0.174
5	Season	0.282	0.253	0.225	0.193
6	Distance	0.311	0.276	0.239	0.201
7	Tree species	0.449	0.417	0.388	0.352

(b)

	df	Bacterial community			Fungal community		
		MS	R^2	Adj. R^2	MS	R^2	Adj. R^2
Replicate	3	0.126	0.064***	0.009	1.808	0.107***	0.055
pH	1	0.844	0.142***	0.092	3.016	0.059***	0.004
OC	1	0.245	0.041***	0.000	0.940	0.019***	0.000
Depth	1	0.103	0.017**	0.000	0.740	0.015**	0.000
Season	1	0.103	0.017**	0.000	1.271	0.025***	0.000
Distance	1	0.170	0.029***	0.000	0.706	0.014**	0.000
Tree species	1	0.821	0.138***	0.088	7.555	0.149***	0.099

Table S3. Multivariate analysis of variance based on weighted UniFrac distances of bacterial and fungal community composition under beech or spruce. (a) Stepwise model testing for multivariate analysis of variance based on weighted UniFrac distances of bacterial and fungal community composition under beech or spruce (64 samples per tree species) with six considered response factors in the order of entering the analysis. Increasing R^2 -values, and adjusted R^2 -values represent chosen model suitability. (b) Analysis of final model with six response variables. R^2 and adjusted R^2 for each variable within the model. Explanatory variables are given in rows in the order of entering the analysis (b). Significant results are indicated by * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Abbreviation: Adj., Adjusted.

(a)

No. of factors	Factor	Beech stand				Spruce stand			
		Bacterial community		Fungal community		Bacterial community		Fungal community	
		R^2	Adj. R^2	R^2	Adj. R^2	R^2	Adj. R^2	R^2	Adj. R^2
1	Replicate	0.063	0.047	0.233	0.221	0.171	0.158	0.323	0.312
2	pH	0.234	0.209	0.265	0.241	0.281	0.257	0.370	0.350
3	OC	0.330	0.297	0.305	0.271	0.300	0.265	0.394	0.363
4	Depth	0.343	0.298	0.319	0.272	0.329	0.283	0.406	0.366
5	Season	0.403	0.352	0.340	0.283	0.351	0.295	0.426	0.377
6	Distance	0.424	0.363	0.360	0.293	0.425	0.364	0.459	0.402

(b)

	Beech stand				Spruce stand			
	Bacterial community		Fungal community		Bacterial community		Fungal community	
	R^2	Adj. R^2	R^2	Adj. R^2	R^2	Adj. R^2	R^2	Adj. R^2
Tree replicate	0.063**	0.000	0.233***	0.152	0.171***	0.084	0.323***	0.252
pH	0.171***	0.084	0.032**	0.000	0.109***	0.015	0.048***	0.000
OC	0.097***	0.002	0.04***	0.000	0.020	0.000	0.023**	0.000
Depth	0.013	0.000	0.013	0.000	0.028*	0.000	0.013	0.000
Season	0.06***	0.000	0.021	0.000	0.022	0.000	0.02*	0.000
Distance	0.021*	0.000	0.020	0.000	0.074***	0.000	0.032***	0.000

Table S4. Analysis of covariance to test the impact of tree replicate, depth, season and distance from the tree trunk on soil chemical parameters under trees. Samples collected under beech (a) and spruce (b) were analyzed. Significant results are indicated by *P < 0.05, **P < 0.01, ***P < 0.001. In this table degrees of freedom (df), mean squares (MS), and F-values are presented.

(a)

	<i>df</i>	pH		OC		N		C:N		Clay	
		<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>
Tree replicate	3	0.03	2.8*	0.01	0.1	0.00	0.1	0.12	0.4	3351	13.7***
Depth	1	0.68	55.7***	22.15	230.7***	0.12	233.2***	17.67	63.2***	12324	50.5***
Season	1	0.01	0.5	0.46	4.8*	0.00	2.9	1.35	4.8*	2680	11.0**
Distance	1	0.02	1.2	3.11	32.4***	0.02	30.8***	3.48	12.5***	10226	41.9***
Residuals	57	0.01		0.10		0.00		0.28		244	

(b)

	<i>df</i>	pH		OC		N		C:N		Clay	
		<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>	<i>MS</i>	<i>F</i>
Tree replicate	3	18.96	93.0***	5.02	7.4***	0.02	10.7***	3.67	2.9*	84533	30.2***
Depth	1	12.87	63.1***	49.23	72.6***	0.11	56.3***	226.48	181.1***	1133	0.4
Season	1	1.65	8.1**	0.53	0.8	0.00	0.1	15.47	12.4***	308	0.1
Distance	1	1.61	7.9**	2.59	3.8	0.00	0.6	16.63	13.3***	11888	4.2*
Residuals	57	0.20		0.68		0.00		1.25		2800	

Table S5. Results of indicator species analysis showing potential indicative OTUs for tree species, sampling depths, sampling distances and sampling season under trees. Bacterial communities under beech (a) and spruce (b) as well as fungal communities under beech (c) and spruce (d) were analyzed. Statistical values: A - species only occurs in sites of this group, B - all sites of this group harbor this species, cutoff (A+B) = 0.7, P - significance value. Relative abundances of OTUs across all 128 samples are depicted. Abbreviations: Rel. abundance, Relative abundance; Acidob., Acidobacteria.

(a)

Indicator group	Bacterial OTU	A	B	P	Rel. abundance (%)
Tree species					
Beech	Xanthomonadales	0.99431	1	0.005	0.216
	Rhodospirillales	0.99403	1	0.005	0.257
	Acidob. subgroup 1	0.99449	0.98438	0.005	0.112
	Acidob. subgroup 2	0.98855	0.98438	0.005	0.081
	Xanthomonadales	0.9919	0.96875	0.005	0.076
	Acidob. subgroup 3 (<i>Bryobacter</i>)	0.95113	1	0.005	0.422
	Gemmatimonadaceae	0.96051	0.98438	0.005	0.241
	Acidob. subgroup 2	0.97674	0.95312	0.005	0.132
	Acidob. subgroup 3 (<i>Bryobacter</i>)	0.93072	1	0.005	0.480
	Acidob. subgroup 2	0.98697	0.9375	0.005	0.094
	Acidob. subgroup 2	0.98664	0.9375	0.005	0.322
	Acidob. subgroup 2	0.90485	1	0.005	0.818
	<i>Gemmatimonas</i>	0.93322	0.96875	0.005	0.184

Table S5a (continued)

Indicator group	Bacterial OTU	A	B	P	Rel. abundance (%)
Depths					
10-20 cm	Chloroflexi	0.8966	0.8125	0.005	0.027
	Acidob. subgroup 1	0.8226	0.7812	0.005	0.019
	Nitrospirales	0.7877	0.8125	0.005	0.045
Distances					
0.5 + 1.5 + 2.5 m	<i>Gemmatimonas</i>	0.8125	0.7083	0.005	0.044
1.5 + 2.5 + 3.5 m	<i>Bryobacter</i>	0.8841	0.8542	0.005	0.042
	Betaproteobacterium	0.86	0.875	0.005	0.062
Season					
Early summer	<i>Bryobacter</i>	0.8686	0.75	0.005	0.042
	<i>Bryobacter</i>	0.8323	0.7188	0.005	0.048
Autumn	<i>Mycobacterium</i>	0.8065	0.9375	0.005	0.048
	<i>Bradyrhizobium</i>	1	0.75	0.005	0.038
	Xanthobacteraceae	0.9478	0.75	0.005	0.035

(b)

Indicator group	Bacterial OTU	A	B	P	Rel. abundance (%)
Tree species					
Spruce	Rhodospirillales	0.9918	0.85938	0.005	0.188
	Acidob. subgroup 1	0.9519	0.82812	0.005	0.332
	Xanthomonadales	0.98683	0.79688	0.005	0.537
	Rhodospirillales	0.93238	0.78125	0.005	0.086
	Acidob. subgroup 1	0.9065	0.79688	0.005	0.151
	Acidob. subgroup 1	0.88484	0.8125	0.005	0.160
	Acidob. subgroup 1	0.95414	0.75	0.005	0.174
	Chloroflexi	0.97448	0.71875	0.005	0.313
	WD272	0.94089	0.73438	0.005	0.062
	Acidob. subgroup 3 (<i>Bryobacter</i>)	0.94694	0.71875	0.005	0.075
Depths					
0-10 cm	Xanthomonadales	0.8868	0.7812	0.005	0.114
	Acetobacteraceae	0.7201	0.8438	0.005	0.082
	Xanthomonadales	0.8293	0.7188	0.005	0.025
	<i>Derrxia</i>	0.8219	0.7188	0.005	0.067

Table S5b (continued)

Indicator group	Bacterial OTU	A	B	P	Rel. abundance (%)
Depths					
10-20 cm	Gaiellales	0.7901	0.8438	0.005	0.100
	Gaiellales	0.8485	0.7812	0.005	0.112
	<i>Caldilinea</i>	0.7536	0.8125	0.005	0.086
	Acidimicrobiales	0.715	0.7812	0.005	0.064
Distances					
2.5 + 3.5 m	Nitrospirales	0.8109	0.7188	0.005	0.192
0.5 + 1.5 + 2.5 m	WD272	0.9549	0.7708	0.005	0.198

(c)

Indicator group	Fungal OTU	A	B	P	Rel. abundance (%)
Tree species					
Beech	<i>Mortierella elongata</i>	0.9987	1	0.005	0.598
	Mortierellomycetes	0.9968	0.9688	0.005	0.253
	Coniochaetaceae	0.9657	0.9219	0.005	0.175
	fungal OTU	0.8301	0.875	0.005	0.160
	<i>Trichoderma</i>	0.9271	0.8594	0.005	0.125
	<i>Xerocomus chrysenteron</i>	1	0.8125	0.005	0.324
	<i>Russula cyanoxantha</i>	0.9987	0.8125	0.005	9.558
	Helotiales	0.9818	0.7656	0.005	0.065
Depths					
0-10 cm	<i>Varicosporium</i>	0.8513	0.7813	0.005	0.053
	<i>Leptodontidium</i>	0.6958	0.75	0.005	0.045
	<i>Volutella</i>	0.8587	0.6875	0.005	0.036
	Didymellaceae	0.8181	0.6875	0.005	0.026
	fungal OTU	0.7895	0.6875	0.005	0.028
Distance					
0.5 m + 1.5 m	<i>Leptodontidium elatius</i>	0.7635	0.6562	0.005	0.025
2.5 m + 3.5 m	<i>Russula cyanoxantha</i>	0.9749	0.5312	0.005	2.140
0.5 m + 1.5 m + 2.5	<i>Piloderma</i>	0.9398	0.5208	0.005	0.091
1.5 m + 2.5 m + 3.5 m	<i>Leotia</i>	0.9179	0.6667	0.005	0.641
	<i>Inocybe</i>	0.9474	0.5208	0.005	1.504

Table S5c (continued)

Indicator group	Fungal OTU	A	B	P	Rel. abundance (%)
Distance					
0.5 m + 1.5 m + 2.5 m + 3.5 m	<i>Mortierella elongata</i>	0.9987	1	0.005	0.598
	Mortierellomycetes	0.9968	0.9688	0.005	0.253
	Coniochaetaceae	0.9657	0.9219	0.005	0.175
	<i>Xerocomus chrysenteron</i>	0.8301	0.875	0.005	0.324
	<i>Russula cyanoxantha</i>	0.9271	0.8594	0.005	9.558
	<i>Trichoderma</i>	1	0.8125	0.005	0.125
	Helotiales	0.9987	0.8125	0.005	0.065
	fungal OTU	0.9818	0.7656	0.005	0.160
Season					
Early summer	<i>Russula cyanoxantha</i>	0.9979	0.5	0.005	0.186
Autumn	Helotiales	0.9936	1	0.005	0.368
	Mortierellaceae	0.7546	0.9688	0.005	0.094
	<i>Mycosphaerella</i>	1	0.7188	0.005	0.027

(d)

Indicator group	Fungal OTU	A	B	P	Rel. abundance (%)
Tree species					
Spruce	<i>Exophiala</i>	0.989	0.9688	0.005	0.178
	<i>Penicillium inflatum</i>	0.9091	0.9688	0.005	0.095
	<i>Hygrophorus discoideus</i>	0.9995	0.7969	0.005	3.169
	Mortierellales	1	0.7813	0.005	0.087
	<i>Hygrophorus pustulatus</i>	0.9997	0.7656	0.005	2.338
	Helotiales	0.9981	0.7656	0.005	0.203
	<i>Penicillium</i>	1	0.75	0.005	0.131
	<i>Amphinema</i>	0.9986	0.7188	0.005	1.387
Depths					
0-10 cm	Helotiales	0.8581	0.9375	0.005	0.203
	<i>Mortierella</i>	0.7384	0.7813	0.005	0.066
10-20 cm	<i>Cryptococcus fuscescens</i>	0.7656	0.7813	0.005	0.102
	<i>Mortierella</i>	0.8424	0.75	0.005	0.068
	<i>Leohumicola</i>	0.9557	0.5313	0.005	0.018
Distance					
0.5 m + 1.5 m + 2.5	<i>Russula</i>	0.9866	0.7292	0.005	4.941
0.5 m + 1.5 m + 2.5 m + 3.5 m	<i>Exophiala</i>	0.989	0.9688	0.005	0.178
	<i>Penicillium inflatum</i>	0.9091	0.8125	0.005	0.095
	<i>Hygrophorus discoideus</i>	0.9995	0.7969	0.005	3.169
	Mortierellales	1	0.7812	0.005	0.087
	Helotiales	0.7812	0.7812	0.005	0.203
	<i>Penicillium</i>	1	0.75	0.005	0.131
	<i>Amphinema</i>	0.9986	0.7188	0.005	1.387
Season					
Early summer	Mortierellales	0.9837	0.625	0.005	0.024
	<i>Amphinema</i>	1	0.5625	0.005	0.168

Table S6. Air temperature, soil temperature, and soil water content data. Five year averages (2008-2012) are presented.

Time	Air temperature (2 m height)		Soil temperature (10 cm depth)		Volumetric water content (10 cm depth)	
	Beech forest	Spruce forest	Beech forest	Spruce forest	Beech forest	Spruce forest
	(°C)		(°C)		(%)	
May	12.1	12.7	10.1	10.6	28.0	23.5
November	4.1	4.2	6.5	7.0	31.5	25.2
January-December	7.3	7.7	8.0	8.4	29.3	25.4

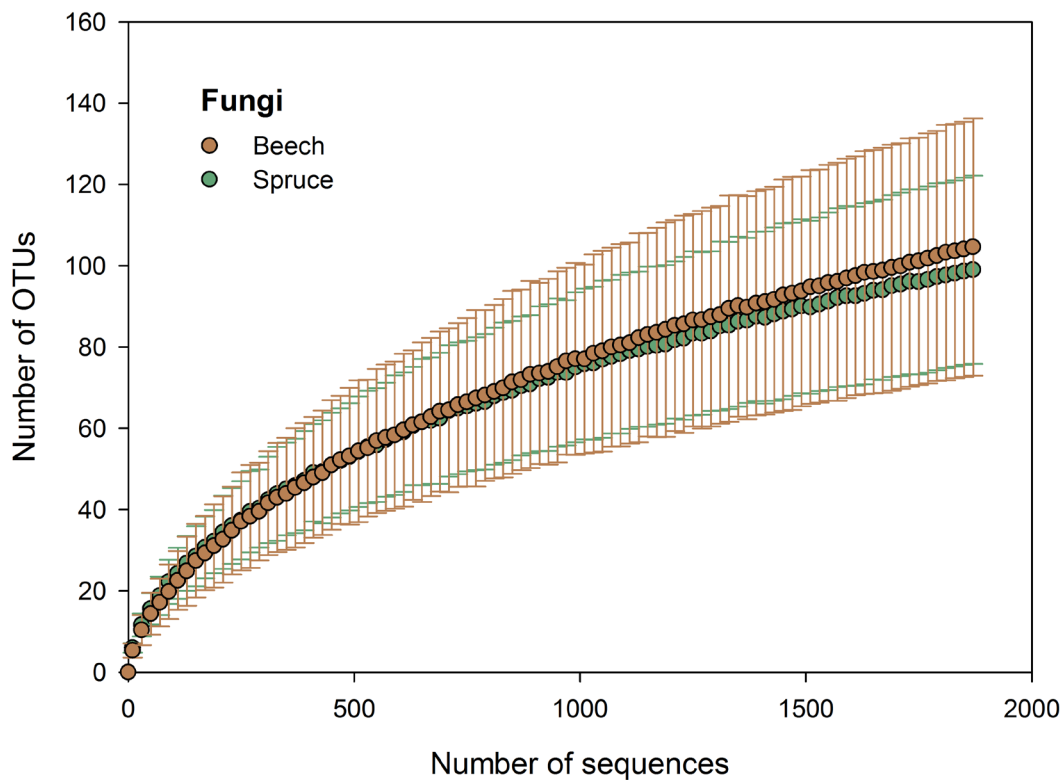
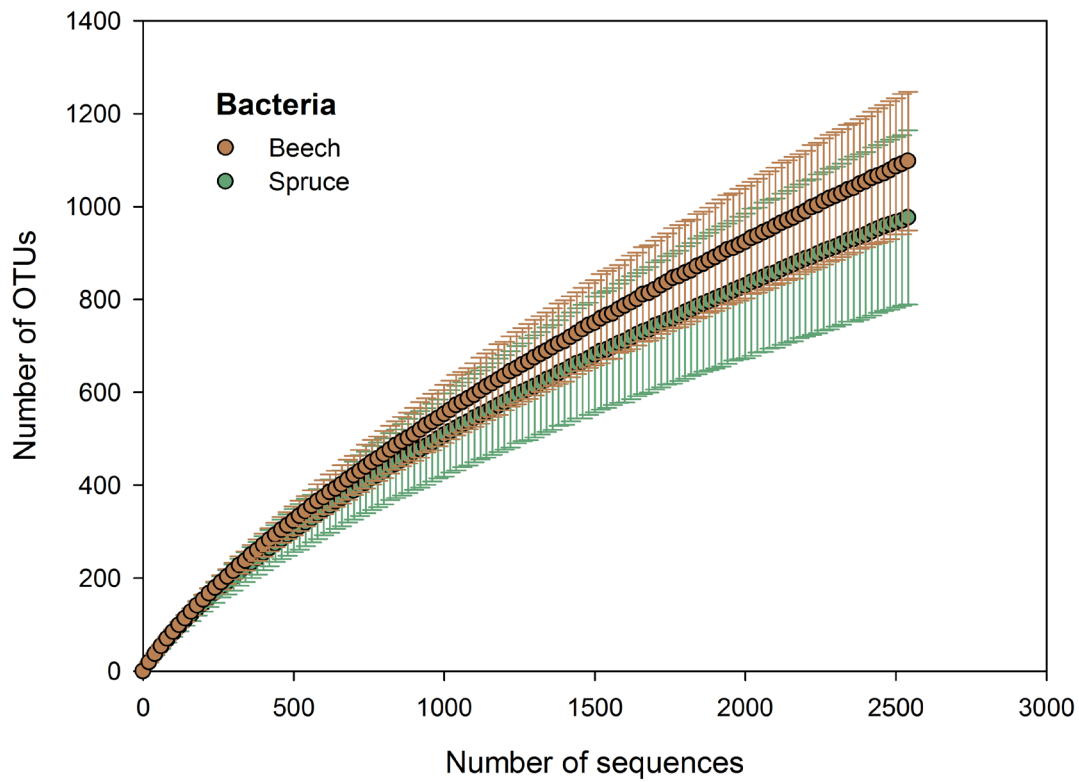


Figure S1. Rarefaction curves indicating the observed numbers of OTUs at a genetic distance of 3%. Samples derived from soil surrounding beech and samples collected under spruce are represented by brown and green color, respectively.

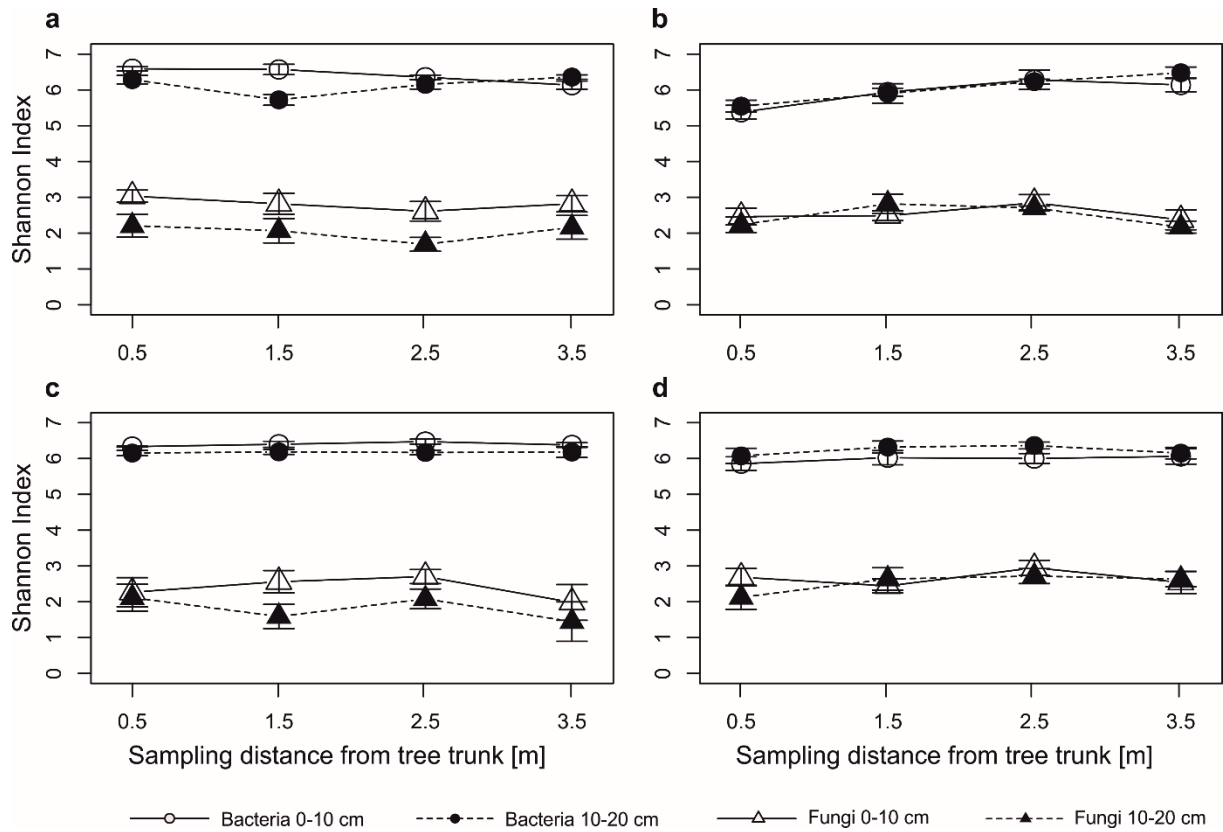


Figure S2. Line plots showing soil bacterial and fungal diversity as assessed by Shannon index at 3% genetic distance under beech in (A) early summer and (C) autumn, and under spruce in (B) early summer and (D) autumn.

8.2.2.***Genome Sequence of the Acetogenic Bacterium Moorella mulderi DSM 14980T*****Genis Andrés Castillo-Villamizar¹, Anja Poehlein¹**

Genome Announc. (2016), 4, e00444-00416

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen

Author contributions to the work:

Performed the experiments: A.P.

Analyzed data: GAC, AP.

Wrote the paper: GAC, AP.

Conceived and designed the experiments: AP.

Genome Sequence of the Acetogenic Bacterium *Moorella mulderi* DSM 14980^T

Genis Andrés Castillo Villamizar,  Anja Poehlein

Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Georg-August University Göttingen, Göttingen, Germany

Here, we report the draft genome sequence of *Moorella mulderi* DSM 14980^T, a thermophilic acetogenic bacterium, which is able to grow autotrophically on H₂ plus CO₂ using the Wood-Ljungdahl pathway. The genome consists of a circular chromosome (2.99 Mb).

Received 11 April 2016 Accepted 13 April 2016 Published 26 May 2016

Citation Castillo Villamizar GA, Poehlein A. 2016. Genome sequence of the acetogenic bacterium *Moorella mulderi* DSM 14980^T. *Genome Announc* 4(3):e00444-16. doi:10.1128/genomeA.00444-16.

Copyright © 2016 Castillo Villamizar and Poehlein. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Anja Poehlein, apoehle3@gwdg.de.

The reduction of CO₂ mediated by acetogenic microorganisms is gaining more interest as a valuable tool for the generation of renewable energy and value-added chemicals (1–3). Thus, homoacetogenic bacteria that use the Wood-Ljungdahl pathway for the CO₂ fixation process have proven to be a main component in this research field (3–8). Among the numerous species of homoacetogens, three organisms have been relatively well studied (*Moorella thermoacetica*, *Acetobacterium woodii*, and *Clostridium ljungdahlii*) (9–13). However, several relevant species remain poorly studied, and the genetic information of many of them remains almost nonexistent or is very limited. Therefore, in this study, we report the draft genome sequence of *Moorella mulderi* DSM 14980^T a thermophilic homoacetogenic anaerobic bacterium originally isolated from a bioreactor with methanol as the energy source (14). Similar to *M. thermoacetica*, *M. mulderi* DSM 14980^T is able to grow on several substrates, including methanol, H₂-CO₂, pyruvate, and glucose. However, several differences have been reported. The optimal temperature of *M. mulderi* DSM 14980^T (65°C) is higher than the optimal temperature reported for *M. thermoacetica* (55 to 60°C). Moreover, in contrast to *M. thermoacetica*, *M. mulderi* DSM 14980^T is able to grow on lactate but cannot use nitrate as an electron acceptor (14).

The MasterPure complete DNA purification kit (Epicentre, Madison, WI, USA) was used to isolate the chromosomal DNA of *M. mulderi* DSM 14980^T. Isolated DNA was used to generate Illumina shotgun sequencing libraries. Sequencing was performed by employing a MiSeq system using MiSeq reagent kit version 3 (600 cycles), as recommended by the manufacturer (Illumina, San Diego, CA, USA), resulting in 2,785,408 paired-end reads (300 bp) that were trimmed using Trimmomatic 0.32 (15). *De novo* assembly performed with the SPAdes genome assembler software version 3.6.2 (16) resulted in 72 contigs (>500 bp) and an average coverage of 188.5-fold.

The genome of *M. mulderi* DSM 14980^T probably consists of a circular chromosome of (2.99 Mb) with an overall G+C content of 53.32%. Gene prediction and annotation were performed using Rapid Prokaryotic Genome Annotation (Prokka) (17). The genome harbored 3 rRNA genes, 52 tRNA genes, 2,240 protein-

coding genes with predicted functions, and 859 genes coding for hypothetical proteins. The cluster of genes encoding enzymes of the methyl and carbonyl branches of the Wood-Ljungdahl pathway is conserved within acetogenic bacteria (18). Therefore, *M. mulderi* DSM 14980^T shows an arrangement identical to the pattern previously identified in *M. thermoacetica* strains ATCC 39073 and DSM 521^T (10, 18). The cluster is composed of eight genes (*acsFABCV*, *cooC*, and *acsDE*) encoding the subunits of the CO dehydrogenase-acetyl-coenzyme A (CoA) synthase complex. The genes encoding the two subunits of the methylene-THF reductase (*metVF*) are located four genes downstream of this cluster.

The genome analysis revealed that *M. mulderi* DSM 14980^T has a bigger genome size than *M. thermoacetica* DSM 521^T (2.52 Mb) and *M. thermoacetica* DSM 2955^T (2.62 Mb) (10, 19).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [LTBC00000000](https://www.ncbi.nlm.nih.gov/nuclink/LTBC00000000). The version described in this paper is version [LTBC01000000](https://www.ncbi.nlm.nih.gov/nuclink/LTBC01000000).

ACKNOWLEDGMENTS

We thank the DAAD for funding and Kathleen Gollnow for technical support.

The funder (DAAD) had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

FUNDING INFORMATION

This work, including the efforts of Genis Andrés Castillo Villamizar, was funded by German Academic Exchange Service Bonn Office (DAAD Bonn).

REFERENCES

- Alissandratos A, Easton CJ. 2015. Biocatalysis for the application of CO₂ as a chemical feedstock. *Beilstein J Org Chem* 11:2370–2387. [http://dx.doi.org/10.3762/bjoc.11.259](https://doi.org/10.3762/bjoc.11.259).
- Drake HL, Gößner AS, Daniel SL. 2008. Old acetogens, new light. *Ann NY Acad Sci* 1125:100–128. [http://dx.doi.org/10.1196/annals.1419.016](https://doi.org/10.1196/annals.1419.016).
- Hawkins AS, McTernan PM, Lian H, Kelly RM, Adams MWW. 2013. Biological conversion of carbon dioxide and hydrogen into liquid fuels

- and industrial chemicals. *Curr Opin Biotechnol* 24:376–384. <http://dx.doi.org/10.1016/j.copbio.2013.02.017>.
4. Bertsch J, Müller V. 2015. Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol Biofuels* 8:210. <http://dx.doi.org/10.1186/s13068-015-0393-x>.
 5. Tremblay P-L, Höglund D, Koza A, Bonde I, Zhang T. 2015. Adaptation of the autotrophic acetogen *Sporomusa ovata* to methanol accelerates the conversion of CO₂ to organic products. *Sci Rep* 5:16168. <http://dx.doi.org/10.1038/srep16168>.
 6. Fontaine FE, Peterson WH, McCoy E, Johnson MJ, Ritter GJ. 1942. A new type of glucose fermentation by *Clostridium thermoaceticum*. *J Bacteriol* 43:701–715.
 7. Andreesen JR, Schaupp A, Neurauter C, Brown A, Ljungdahl LG. 1973. Fermentation of glucose, fructose, and xylose by *Clostridium thermoaceticum*: effect of metals on growth yield, enzymes, and the synthesis of acetate from CO₂. *J Bacteriol* 114:743–751.
 8. Schaupp A, Ljungdahl LG. 1974. Purification and properties of acetate kinase from *Clostridium thermoaceticum*. *Arch Microbiol* 100:121–129. <http://dx.doi.org/10.1007/BF00446312>.
 9. Pierce E, Xie G, Barabote RD, Saunders E, Han CS, Detter JC, Richardson P, Brettin TS, Das A, Ljungdahl LG, Ragsdale SW. 2008. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ Microbiol* 10:2550–2573.
 10. Poehlein A, Bengelsdorf FR, Esser C, Schiel-Bengelsdorf B, Daniel R, Dürre P. 2015. Complete genome sequence of the type strain of the acetogenic bacterium *Moorella thermoacetica* DSM 521^T. *Genome Announc* 3(5):e01159-15. <http://dx.doi.org/10.1128/genomeA.01159-15>.
 11. Bertsch J, Müller V. 2015. CO metabolism in the acetogen *Acetobacterium woodii*. *Appl Environ Microbiol* 81:5949–5956. <http://dx.doi.org/10.1128/AEM.01772-15>.
 12. Younesi H, Najafpour G, Mohamed AR. 2005. Ethanol and acetate production from synthesis gas via fermentation processes using anaerobic bacterium, *Clostridium ljungdahlii*. *Biochem Eng J* 27:110–119. <http://dx.doi.org/10.1016/j.bej.2005.08.015>.
 13. Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, Wollherr A, Ehrenreich A, Liebl W, Gottschalk G, Dürre P, Demain AL. 2010. *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc Natl Acad Sci USA* 107:13087–13092. <http://dx.doi.org/10.1073/pnas.1004716107>.
 14. Balk M, Weijma J, Friedrich MW, Stams AJM. 2003. Methanol utilization by a novel thermophilic homoacetogenic bacterium, *Moorella mulderi* sp. nov., isolated from a bioreactor. *Arch Microbiol* 179:315–320.
 15. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
 16. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
 17. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
 18. Poehlein A, Cebulla M, Ilg MM, Bengelsdorf FR, Schiel-Bengelsdorf B, Whited G, Andreesen JR, Gottschalk G, Daniel R, Dürre P. 2015. The complete genome sequence of *Clostridium aceticum*: a missing link between Rnf- and cytochrome-containing autotrophic acetogens. *mBio* 6:e01168-15. <http://dx.doi.org/10.1128/mBio.01168-15>.
 19. Bengelsdorf FR, Poehlein A, Esser C, Schiel-Bengelsdorf B, Daniel R, Dürre P. 2015. Complete genome sequence of the acetogenic bacterium *Moorella thermoacetica* DSM 2955^T. *Genome Announc* 3:e01157-15. <http://dx.doi.org/10.1128/genomeA.01157-15>.

8.2.3.***First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic Sporomusa sphaeroides Strain E (DSM 2875)*****Genis Andrés Castillo-Villamizar¹, Rolf Daniel¹, Anja Poehlein¹**

Genome Announc. (2017), 5 (12), e00037-17

¹Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University, Göttingen

Author contributions to the work:

Performed the experiments: AP.

Analyzed data: GAC, AP.

Wrote the paper: GAC, AP, RD.

Conceived and designed the experiments: AP, RD.



First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic *Sporomusa sphaeroides* Strain E (DSM 2875)

Genis Andrés Castillo Villamizar, Rolf Daniel,  Anja Poehlein

Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Georg-August University Göttingen, Göttingen, Germany

ABSTRACT Here, we report the draft genome sequence of *Sporomusa sphaeroides* strain E (DSM 2875), a strict anaerobic homoacetogenic bacterium. It is able to grow autotrophically on different one-carbon compounds. The strain possesses several genes of the Wood-Ljungdahl pathway. The genome consists of a single chromosome (4.98 Mb).

The autotrophic metabolism of diverse acetogenic bacteria is used for the development of economically relevant chemicals such as acetate, ethanol, butyrate, and butanol. Likewise, the quest for alternative, renewable, and sustainable energy sources resulted in an increased interest for processes involving anaerobic digestion. Among the most studied organisms involved in anaerobic digestion processes are Gram-positive acetogens like *Clostridium ljungdahlii*, *C. aceticum*, and the thermophile *Moorella thermoacetica*. Gram-negative acetogens comprise several species of the *Sporomusa* genus (1–3).

The publication of genome sequences of many acetogens involved in biotechnological processes improves knowledge and drives the development of new and more efficient production platforms (4–8). In this study, we report the draft genome sequence of *Sporomusa sphaeroides* E (DSM 2875). This organism has been isolated from mud samples of the German Leine River (9).

A MasterPure complete DNA purification kit (Epicentre, Madison, WI, USA) was used to isolate chromosomal DNA of *S. sphaeroides* E (DSM 2875). The extracted DNA was used to generate 454 shotgun, 454 paired-end, and Illumina shotgun libraries (paired-end) according to the manufacturers' protocols (Roche Life Sciences, Mannheim, Germany, and Illumina, Inc., San Diego, CA, USA). The libraries were sequenced using a 454 GS-FLX system (Titanium GS70 chemistry; Roche Life Sciences, Mannheim, Germany) and a Genome Analyzer II (Illumina, Inc.). Sequencing resulted in 251,686 454 shotgun reads, 100,698 454 paired-end reads (1.6-kb and 2.8-kb insert sizes), and 7,621,534 Illumina paired-end reads (112 bp). Assembly of the reads using Roche Newbler assembly software 2.6 for scaffolding and MIRA software (10) resulted in 35 scaffolds with 108 contigs. The average coverage was 26.8-fold for 454 and 171.44-fold for Illumina. Some gaps were closed using PCR and Sanger sequencing of the products. Analysis of the obtained sequences was completed using Gap4 (version 4.11) software of the Staden package (11, 12). The final draft genome of *S. sphaeroides* (16 contigs) consists of a single chromosome of 4.98 Mb with an overall G+C content of 47.21%. Gene prediction and annotation were performed using Prokka (13). The draft genome harbored 17 rRNA genes, 88 tRNA genes, 3,564 protein-coding genes with predicted functions, and 1,150 genes coding for hypothetical proteins.

Received 11 January 2017 Accepted 19 January 2017 Published 23 March 2017

Citation Villamizar GAC, Daniel R, Poehlein A. 2017. First insights into the genome sequence of the strictly anaerobic homoacetogenic *Sporomusa sphaeroides* strain E (DSM 2875). *Genome Announc* 5:e00037-17. <https://doi.org/10.1128/genomeA.00037-17>.

Copyright © 2017 Villamizar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Anja Poehlein, apoehle3@gwdg.de.

The cluster of genes encoding enzymes of the methyl and carbonyl branches of the Wood-Ljungdahl pathway is present in the genome and showed the same organization as described for *S. ovata* DSM 2662 (6). The genome contains genes encoding for enzymes involved in the metabolism of one-carbon compounds, including *cooS* for the putative synthesis of the carbon monoxide dehydrogenase (CODH) required for growing on CO. Five genes of the formate dehydrogenase (*fdhs*) complex involved in the oxidation of formate were also detected. Finally, the methyltransferase genes *mtaB* and *mtaD*, required for methanol-specific methyl transfer (14), were also present.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [LSLJ00000000](https://doi.org/10.1111/1462-2920.2008.01679.x). The version described in this paper is the first version, LSLJ01000000.

ACKNOWLEDGMENTS

This work was supported by a grant from the Bundesministerium für Bildung und Forschung (BMBF) within the ERA-IB 7 framework project "OBAC": Overcoming Energetic Barriers in Acetogenic Conversion of Carbon Dioxide. We thank Melanie Heine-mann for technical support.

REFERENCES

1. Nevin KP, Hensley SA, Franks AE, Summers ZM, Ou J, Woodard TL, Snoeyenbos-West OL, Lovley DR. 2011. Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Appl Environ Microbiol* 77:2882–2886. <https://doi.org/10.1128/AEM.02642-10>.
2. Tremblay PL, Zhang T. 2015. Electrifying microbes for the production of chemicals. *Front Microbiol* 6:201. <https://doi.org/10.3389/fmicb.2015.00201>.
3. Hawkins AS, McTernan PM, Lian H, Kelly RM, Adams MWW. 2013. Biological conversion of carbon dioxide and hydrogen into liquid fuels and industrial chemicals. *Curr Opin Biotechnol* 24:376–384. <https://doi.org/10.1016/j.copbio.2013.02.017>.
4. Poehlein A, Gottschalk G, Daniel R. 2013. First insights into the genome of the Gram-negative, endospore-forming organism *Sporomusa ovata* strain H1 DSM 2662. *Genome Announc* 1(5):e00734-13. <https://doi.org/10.1128/genomeA.00734-13>.
5. Castillo Villamizar GA, Poehlein A. 2016. Genome sequence of the acetogenic bacterium *Moorella mulderi* DSM 14980T. *Genome Announc* 4(3):e00444-16. <https://doi.org/10.1128/genomeA.00444-16>.
6. Poehlein A, Cebulla M, Ilg MM, Bengelsdorf FR, Schiel-Bengelsdorf B, Whited G, Andreesen JR, Gottschalk G, Daniel R, Dürre P. 2015. The complete genome sequence of *Clostridium aceticum*: a missing link between Rnf- and cytochrome-containing autotrophic acetogens. *mBio* 6:e01168-15. <https://doi.org/10.1128/mBio.01168-15>.
7. Pierce E, Xie G, Barabote RD, Saunders E, Han CS, Detter JC, Richardson P, Brettin TS, Das A, Ljungdahl LG, Ragsdale SW. 2008. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoacetica*). *Environ Microbiol* 10:2550–2573. <https://doi.org/10.1111/1462-2920.2008.01679.x>.
8. Schiel-Bengelsdorf B, Dürre P. 2012. Pathway engineering and synthetic biology using acetogens. *FEBS Lett* 586:2191–2198. <https://doi.org/10.1016/j.febslet.2012.04.043>.
9. Möller B, Oßmer R, Howard BH, Gottschalk G, Hippe H. 1984. *Sporomusa*, a new genus of Gram-negative anaerobic bacteria including *Sporomusa sphaeroides* spec. nov. and *Sporomusa ovata* spec. nov. *Arch Microbiol* 139:388–396.
10. Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Comput Sci Biol Proc* 99:45–56.
11. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>.
12. Staden R, Beal KF, Bonfield JK. 2000. The Staden package, 1998. *Methods Mol Biol* 132:115–130.
13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
14. Visser M, Pieterse MM, Pinkse MWH, Nijse B, Verhaert PDEM, de Vos WM, Schaap PJ, Stams AJM. 2016. Unravelling the one-carbon metabolism of the acetogen *Sporomusa* strain An4 by genome and proteome analysis. *Environ Microbiol* 18:2843–2855. <https://doi.org/10.1111/1462-2920.12973>.

8.3. Acknowledgments

First of all, I would like to thank Prof. Dr. Rolf Daniel, not just for the opportunity to accomplish one of the biggest goals in my life but also for all the encouraging advice and support over all these years. Thanks also for creating such a pleasant and enjoyable work atmosphere.

I am grateful to Dr. Michael Hoppert for his readiness to be the second examiner of this thesis and because of the evaluation letters and support during my time as a DAAD scholarship holder.

I am particularly and deeply thankful to Dr. Heiko Nacke whose unlimited good advice, support, and patience have been crucial in every single step of this process.

I want to thank Dr. Anja Poehlein, and Dr. Silja Brady who were my first contact to the lab back during my master courses. Their kindness, motivation, and professionalism confirmed to me right away that I was in the best possible place.

I also would like to thank all the outstanding post-doc team (Dr. Dominik Schneider, Dr. Bernd Wemheuer, Dr. Robert Hertel, Dr. Jacqueline Hollensteiner, Dr. Heiko Liesegang, Dr. Birgit Pfeiffer), the scientific coordinator Dr. Petra Ehrenreich and some former members of the department (Dr. Andrea Thürmer and Dr. Sascha Dietrich). For me it was always a privilege being able to count on so many people with such a variety of skills always willing to help and offer knowledge, experience, and advice.

Similarly, I would like to thank our technical staff and our team of technical assistants. In particular Mechthild Bömeke from who I have learned a lot during all these years and whose organization of the lab has made my life at the department much easier.

A gigantic thank you, to my fellow doctoral students, the current cohort: Tati, Dirk, Avi, Ines, Richard, Randi, Stefani, Simone, Mingi Inka, Cynthia, Florian. And the ones who preceded Kristin, Sarah and Amelie. I have enjoyed my time with you guys as you have no idea. I will always keep in my happy memories our parties, dinners, drinking seminars, trips, grills, chats, movies, discussions, laughs etc. I must not forget to thank the booty hearth criminal who gave me tons of energy in form of cookies for many years.

This thesis was made also with the enormous contribution of several students who worked with me. Marc Böhning, Karolin Foerster, Tobias Schilling, Sanghamitra

Narredula (Pandú), Laura Griese and off course my MBL partner Katrina Funkner. Their motivation and effort helped me to reach this goal.

I would also like to thank Dem Deutschen Akademischen Austauschdienst (DAAD), Colciencias and the Corporación para la investigación de la corrosión C.I.C. for the economic support during my time of study in Germany in form of the fellowship program ALECOL. Referring to this my particularly deep thank to Dr. Jorge Hernando Panqueva. His motivation, confidence and full support, have allowed me to accomplish my goals. I am aware of the opportunity he has given me, and I am looking forward to responding positively to the expectations.

I would like to thank specially to the Puentes-Bachmann, Edinson and, Jenny and to Josh not only for their critical reading of some parts of this thesis but also for their unconditional friendship.

This section would not be complete without acknowledging the efforts of my beloved old family in Colombia. Specially to my mother and my grandmother who have been always my biggest fans and supporters. They sacrificed with happiness (I assume) many things for me and my goals. They are my heroes and my role models. I also could not be here without what is my new family Peter, Sabine, and the entire German crew. They have given me tons of love and kindness for years.

Finally, when I left Colombia, I thought that I was going away from my home. For a while I was, but in Germany I got a new home. I met my beloved wife Sina who is my everything. She has supported and helped me intensively during many stages of this thesis. Moreover, with her, we founded a family and now we have our littles, Luc and Mati. They have been living with my thesis stress, but always respond with fun, laugh and love. They gave me tons of calm and joy during the most difficult hours.

8.4. Curriculum Vitae

Genis Andrés Castillo Villamizar

Education

Since 2014	Georg-August-Universität Göttingen Ph.D candidate: Biology (Genomics and Metagenomics) Advisor: Prof. Dr. Rolf Daniel
2011 - 2014	Georg-August-Universität Göttingen Msc. Microbiology and Biochemistry Advisor: Prof. Dr. Rolf Daniel
1997 - 2004	Universidad Industrial de Santander (UIS) B.Sc. Biology Advisor: Med. Clara Ines Vargas

Practical experience

Since 2007 -	Researcher Corporación para la investigación de la Corrosión (C.I.C). Bio-corrosion
2005 – 2008	Teaching Assistant, Biology school, Universidad Industrial de Santander. Population genetics and Molecular Biology
2004 – 2007	Research /Assistant, Molecular Biology and biotechnology laboratory. Universidad Industrial de Santander (UIS)

Honors and awards

- 2011 - 2016 German Academic Exchange Service (DAAD) Msc / Phd scholarship
- 2001 - 2003 Outstanding Student of the Year
Awarded by Universidad Industrial de Santander, School of Biology
-

Publications

- Castillo Villamizar GA**, Nacke H, Boehning M, Herz K, Daniel R. 2019. Functional metagenomics reveals an overlooked diversity and novel features of soil-derived bacterial phosphatases and phytases. *mBio*, 10, e01966-01918.
- Castillo Villamizar GA**, Nacke H, Griese L, Tabernero L, Funkner K, Daniel R. 2019. Characteristics of the first protein tyrosine phosphatase with phytase activity from a soil metagenome. *Genes*, 10, 101.
- Castillo Villamizar GA**, Poehlein A. 2016. Genome sequence of the acetogenic bacterium *Moorella mulderi* dsm 14980t. *Genome Announc*, 4, e00444-00416.
- Nacke H, Goldmann K, Schöning I, Pfeiffer B, Kaiser K, **Castillo-Villamizar GA**, Schrupf M, Buscot F, Daniel R, Wubet T. 2016. Fine spatial scale variation of soil microbial communities under european beech and norway spruce. *Front Microbiol*, 7, 2067.
- Reyes-Reyes MA, Puentes-Cala EA, Casanova-Montes EL, López-Deluche F, Panqueva-Álvarez JH, **Castillo-Villamizar GA**. 2018. Inmovilización de bacterias potencialmente degradadoras de petróleo crudo en matrices orgánicas naturales y sintéticas. *Rev Int Contam Ambie*, 34, 597-609.
- Villamizar GAC**, Daniel R, Poehlein A. 2017. First Insights into the Genome Sequence of the Strictly Anaerobic Homoacetogenic *Sporomusa sphaeroides* Strain E (DSM 2875). *Genome Announc*, 5.
- Villamizar GAC**, Nacke H, Daniel R. 2016. Function-based metagenomic library screening and heterologous expression strategy for genes encoding phosphatase activity., p 249-260. *Methods Mol Biol.* (2017), 1539:249-260.

Hernandez-Torres J, **Villamizar GAC**, Salgar-Chaparro SJ, Silva-Plata BA, Daza ODS, Martinez-Perez F, Umana RM, Levy M, Fuentes JL. 2016. Prokaryotic community characterization in a mesothermic and water-flooded oil reservoir in Colombia. *Geomicrobiol J*, 33, 110-117.

Niño EDV, Garnica H, Dugar-Zhabon V, **Castillo G**. 2014. Behavior of AISI SAE 1020 steel implanted by titanium and exposed to bacteria sulphate deoxidizer. *J Phys Conf Ser*, 511, 012080.

Serrano-Serrano ML, Hernandez-Torres J, **Castillo-Villamizar G**, Debouck DG, Sanchez MI. 2010. Gene pools in wild Lima bean (*Phaseolus lunatus* L.) from the Americas: evidences for an Andean origin and past migrations. *Mol Phylogenet Evol*, 54, 76-87.

Motta-Aldana JR, Serrano-Serrano ML, Hernández-Torres J, **Castillo-Villamizar G**, Debouck DG, Chacón S MI. 2010. Multiple origins of lima bean landraces in the Americas: Evidence from chloroplast and nuclear DNA polymorphisms. *Crop Sci*, 50, 1773.

Suarez GA, **Castillo G**, Chacon SM. 2008. Genetic diversity and spatial genetic structure within a population of an aromatic shrub, *Lippia origanoides* (Verbenaceae), in the Chicamocha Canyon, northeastern Colombia. *Genet Res*, 90, 455-465.

Suárez González AR, Martínez Nuñez FO, **Castillo Villamizar GA**, Chacón S MI. 2007. Molecular characterization of aromatic species of the genus *lippia* from the Colombian neotropics. *Acta Horti* doi:10.17660/ActaHortic.2007.756.14, 129-138.