DISS. ETH NO. 23284

DEVELOPMENT AND APPLICATION OF MOLECULAR TOOLS TO INVESTIGATE MICROBIAL ALKALINE PHOSPHATASE GENES IN SOIL

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SABINE ANNE RAGOT

Master of Science UZH in Biology

born on 25.02.1987

citizen of Fribourg, FR

accepted on the recommendation of

Prof. Dr. Emmanuel Frossard, examiner

PD Dr. Else Katrin Bünemann-König, co-examiner

Prof. Dr. Michael Kertesz, co-examiner

Dr. Claude Plassard, co-examiner

2016

ABSTRACT

Phosphatase enzymes play an important role in soil phosphorus cycling by hydrolyzing organic phosphorus to orthophosphate, which can be taken up by plants and microorganisms. PhoD and PhoX alkaline phosphatases and AcpA acid phosphatase are produced by microorganisms in response to phosphorus limitation in the environment. In this thesis, the current knowledge of the prevalence of *phoD* and *phoX* in the environment and of their taxonomic distribution was assessed, and new molecular tools were developed to target the *phoD* and *phoX* alkaline phosphatase genes in soil microorganisms. The newly-designed primers were then used to identify *phoD*- and *phoX*-harboring microorganisms and to explore the relationships between selected environmental factors and the *phoD*- and *phoX*-harboring community composition and structure in thirty sites across three land-uses and six soil types in Australia and Switzerland. Additionally, the total and active *phoD*-harboring community composition and structure as affected by phosphate depletion and pH were studied in a long-term fertilization trial in grassland characterized by a natural pH gradient on the site.

The newly-designed primers amplified *phoD* and *phoX* in soil microorganisms with a good coverage and specificity. The *phoD* gene was found in 1 archaeal, 13 bacterial and 2 fungal phyla, and the *phoX* gene in 1 archaeal and 16 bacterial phyla. Dominant *phoD*-harboring phyla were *Actinobacteria, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Gemmatimonadetes, Planctomycetes* and *Proteobacteria*, while abundant *phoX*-harboring phyla were *Acidobacteria, Actinobacteria, Chloroflexi, Planctomycetes, Proteobacteria* and *Verrucomicrobia*. The composition and structure of *phoD*- and *phoX*-harboring communities were significantly correlated with climate, soil group, land-use, pH and soil nutrient concentrations. However, the significance of these correlations differed in

intensity between the *phoD*- and the *phoX*-harboring community. Additionally, our case study in the long-term fertilization trial showed that the total and active *phoD*-harboring communities were on the whole composed of similar phyla with, however, different relative abundances. Furthermore, phosphate depletion affected the composition and structure of the active *phoD*-harboring community only, while soil pH impacted the composition and structure of both the total and active *phoD*-harboring community.

Additionally, primers targeting the *acpA* acid phosphatase gene were designed and preliminary work was conducted on the same thirty sites used to study *phoD* and *phoX*. Our results suggest that *acpA* is not as widespread in bacteria as *phoD* and *phoX*. Dominant *acpA*-harboring orders included *Burkholderiales, Methylococcales, Pseudomonadales* and *Rhizobiales*.

The primers designed in this thesis revealed a large diversity of *phoD* and *phoX* in soil and represent valuable tools to study *phoD*- and *phoX*-harboring communities in environmental samples. Despite differences in environmental factors, dominant *phoD*-harboring phyla were generally similar in all samples, while dominant *phoX*-harboring phyla differed substantially between the samples. Nonetheless, our results suggest that the composition and structure of the *phoD*- and the *phoX*-harboring communities are affected by the same environmental factors.

RÉSUMÉ

Les phosphatases jouent un rôle capital dans le cycle du phosphore dans le sol. En tant qu'enzymes, elles catalysent l'hydrolyse des composés organiques phosphatés en orthophosphate, l'unique forme de phosphore disponible pour les plantes et les microorganismes du sol. Les phosphatases alcalines PhoD et PhoX et la phosphatase acidique AcpA sont sécrétées par des microorganismes. Elles hydrolysent essentiellement les phosphomonoesters qui représentent souvent le type de phosphore organique le plus abondant dans le sol. Dans ce travail de thèse, de nouveaux outils moléculaires ciblant les gènes des phosphatases alcalines *phoD* et *phoX* dans les microorganismes du sol ont été développés. Ces outils ont par la suite été utilisés dans trente sols classés en cinq groupes différent et incluant trois types d'utilisation du sol en Australie et en Suisse, afin d'identifier les microorganismes clés ayant les gènes *phoD* et *phoX* et d'explorer les liens entre la structure et la composition de la communauté possédant ces genes et les éléments environnementaux. De plus, un essai de fertilisation phosphatée de longue durée dans une prairie caractérisée par un gradient de pH naturel a permis l'étude des effets du l'appauvrissement en phosphate et du pH sur la composition et la structure des communautés active et totale ayant le gène *phoD*.

Ces nouvelles amorces ont permis d'amplifier les gènes *phoD* et *phoX* chez les microorganismes du sol avec une bonne spécificité et une bonne couverture de la diversité. Dans notre étude, le gène *phoD* était présent dans un phylum archéal, treize phyla bactériens et deux phyla fongiques, tandis que le gène *phoX* était présent dans un phylum archéal et seize phyla bactériens. Les phyla clés ayant le gène *phoD* incluaient des *Actinobacteria, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Gemmatimonadetes, Planctomycetes* et *Proteobacteria*, tandis que les phyla clés ayant le gène *phoX* incluaient des *Acidobacteria, Actinobacteria, Chloroflexi, Planctomycetes, Proteobacteria* et *Verru-*

comicrobia. La composition et la structure des communautés ayant les gènes *phoD* et *phoX* étaient significativement corrélées au climat, groupe de sol, type d'utilisation du sol, pH et concentrations de nutriments du sol. L'étude de cas dans l'essai de fertilisation phosphatée a montré que les communautés active et totale ayant le gène *phoD* étaient composées des mêmes phyla. De plus, alors que le pH influençait la composition et la structure des communautés active et totale possédant le gène *phoD*, l'appauvrissement en phosphate affectait uniquement la communauté active.

De plus, des amorces amplifiant le gène *acpA* ont été développées et des étude préliminaires sur la prévalence et la composition du gène *acpA* dans les trente sols utilisés pour étudier les gènes *phoD* et *phoX* ont été menées. Nos résultats suggèrent que le gène *acpA* n'est pas aussi répandu dans les bactéries que les gènes *phoD* et *phoX*. Les ordres clés comprenant le gène *acpA* incluaient les *Burkholderiales, Methylococcales, Pseudomonadales* and les *Rhizobiales*.

Les amorces développées dans cette thèse révèlent la grande diversité des microorganismes du sol ayant les gènes *phoD* et *phoX*. En dépit des différences des éléments environnementaux, les phyla ayant le gène *phoD* étaient similaires, tandis que les phyla ayant le gène *phoX* variaient considérablement. Toutefois, nos résultats suggèrent que la composition et la structure des communautés ayant les gènes *phoD* et *phoX* sont influencées par les mêmes éléments environnementaux. Les nouvelles amorces présentées de ce travail sont de précieux outils pour étudier les communautés possédant les gènes *phoD* et *phoX* dans l'environment.

ABBREVIATIONS

ADONIS	Analysis of Variance Using Distance Matrices
ANOSIM	Analysis of Dissimilarity
ANOVA	Analysis of Variance
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
cAMP	Cyclic Adenosine Monophosphate
BLAST	Basic Local Alignment Search Tool
C	Carbon
$CaCO_3$	Carbonate
CD database	Conserved Domain Database
C_{mic}	Microbial Carbon
COG	Clusters of Orthologous Groups
DNA	Deoxyribonucleic Acid
ENA	European Nucleotide Archive
IMG/M database	Integrative Microbial Genomes and Metagenomes Database
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
N	Nitrogen
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NMDS	Non-Metric Dimensional Scaling
N_{mic}	Microbial Nitrogen
P	Phosphorus
PCR	Polymerase Chain Reaction
P_{res}	Resin-extractable Phosphorus
P_{org}	Organic Phosphorus
qPCR	Quantitative Real-Time PCR
RDA	Redundancy Analysis
RNA	Ribonucleic Acid
SFF	Standard Flowgram Format
S_{est}	Estimated Species Richness
S_{obs}	Observed Species Richness
TAT	Twin Arginine Translocation
TC	Total Carbon
TOC	Total Organic Carbon
TP	Total Phosphorus
TN	Total Nitrogen
T-RFLP	Terminal-Restriction Fragment Length Polymorphism

CONTENTS

1

GENERAL INTRODUCTION

1.1 Phosphorus in soil

As an essential macronutrient, phosphorus (P) is involved in fundamental biochemical reactions including genetic inheritance and expression (DNA, RNA), energy storage and transfer (ATP), intracellular signaling (cAMP), and structural integrity provided by membranes (phospholipids, teichoic acids, lipopolysaccharides) and bones (hydroxyapatite) (Westheimer 1987).

P is the 11th most abundant element in the Earth's crust and is primarily found as apatite. P-bearing bedrock represents the principal source of P in soil. It is released from the bedrock through weathering processes and is then found as orthophosphate anions (PO_4^{3-}, HPO_4^{2-}, $H_2PO_4^-$), non-occluded (sorbed onto Ca^{2+}-, $Fe^{2+/3+}$- and Al^{3+}-oxides) and occluded P-bearing minerals (Ca^{2+}-, $Fe^{2+/3+}$- and Al^{3+}-phosphate) (Yang and Post 2011). At the initial stage of soil development, orthophosphate anions are promptly assimilated by pioneer microorganisms and plants, which start building up the soil organic P fraction (Tamburini et al. 2012). At the mid stage, the reservoir of primary P minerals is diminished and the secondary P minerals and organic P fractions increase. At the late stage, the inorganic P fraction is principally composed of non-occluded and occluded P minerals and the organic P fraction stabilizes (Walker and Syers 1976).

Soil P forms can be characterized after their chemical characteristics as inorganic and organic pools and additionally described as labile and stable pools, referring to their degree of physicochemical and biological stability (Figure 1.1). Total P concentration in soil ranges from 100 to 3000 μg g^{-1} (Condron et al. 2005). The organic P fraction generally represents 20 to 65% of total P, ranging from as low as 4% to as high as 95% in certain soils (Harrison 1987). It consists of three pools (i) microbial, (ii) labile and (iii) stable organic P (Figure 1.1). Each pool consists of a variety of organic P compounds which are generally defined after the type of chemical bonds between P and the C moiety. These include orthophosphate monoesters (inositol phosphates, phosphoproteins, mononucleotides, sugar phosphates) and –diesters (nucleic acids, phospholipids, teichoic

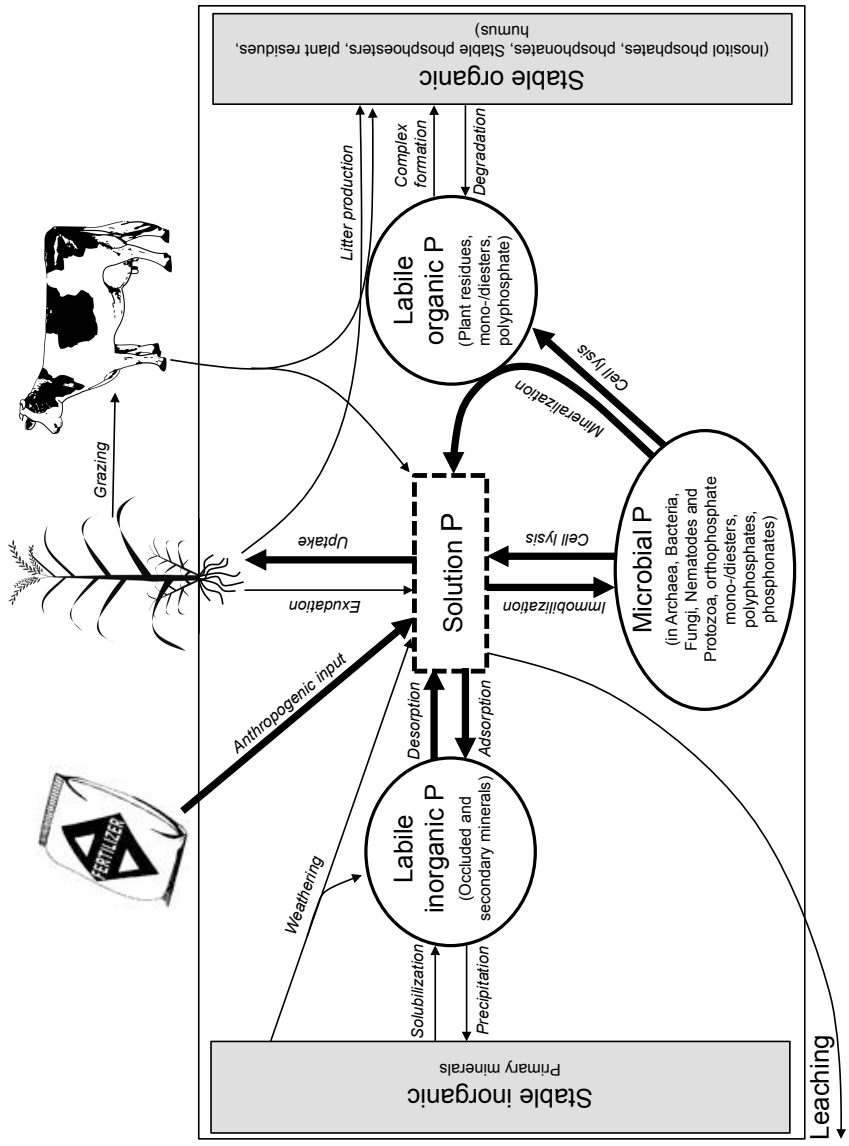


Figure 1.1. Conceptual model of soil P cycle modified (adapted from Plante (2007).

acids) and phosphonates (Condron et al. 2005). Microbial P is often described as a highly dynamic P pool due to the efficient immobilization of P by microorganisms ranging between 0.3 and 5.5 mg P kg^{-1} d^{-1} (Bünemann et al. 2012), and its fast turnover time ranging between 20 days and 1.25 years for grassland soils (Oberson and Joner 2005; Tamburini et al. 2012). As a result, P is generally cycled through microorganisms before being released into the soil solution (Tamburini et al. 2012). Microbial P can represent between 0.4 and 2.5% of total P in cropped soils and up to 7.5% in grassland soils (Oberson and Joner 2005).

The inorganic P fraction typically accounts for 25 to 80% of the total P in soil (Harrison 1987). The inorganic orthophosphate present in soluble form in the soil solution is the only readily available P form for plants and soil microorganisms, commonly representing less than 1% of total P (Harrison 1987). The low P availability in soil is due to the physicochemical properties of orthophosphate, which is highly negatively charged and has a weak hydration shell of one or two molecules of water (Blades et al. 1996), resulting in strong adsorption principally onto Ca^{2+}-, $Fe^{2+/3+}$- and Al^{3+}-oxides and clay mineral surfaces as well as precipitation as Ca^{2+}-, $Fe^{2+/3+}$- and Al^{3+}-phosphates in soil (Frossard et al. 1995). As a consequence, P is a common growth-limiting factor in many ecosystems (Vitousek et al. 2010), forcing terrestrial organisms to develop mechanisms to acquire and recycle P. For example, plants and microorganisms can increase the number of high affinity P transporters (*pst*) in the cytoplasmic membrane to efficiently assimilate P (Hermans et al. 2006; Vershinina and Znamenskaya 2002; Plassard and Dell 2010) and build up P storage (e.g. polyphosphates) (Nicholls and Osborn 1979). Moreover, plants can change the rhizosphere conditions to enhance P solubility by secreting exudates (e.g. organic anions) that can chelate orthophosphate and/or reduce pH (Richardson and Simpson 2011; Plassard et al. 2015). Plants can improve their volume to mass root ratio by increasing root length and root hair density (Schachtman et al. 1998). They can also foster symbiosis with arbuscular mycorrhizal fungi that can extend the surface area for nutrient uptake or with ectomycorrhizal fungi that can solubilize inorganic P

and hydrolyze organic P (Schachtman et al. 1998; Plassard et al. 2011). Plants and soil organisms have evolved to produce a variety of enzymes called phosphatases that hydrolyze organic P in soil (Richardson and Simpson 2011; Plassard et al. 2015).

1.2 Phosphatases

1.2.1 Phosphatase classes

Organic P represents an important source of P for plants and microorganisms. Generally, organic P compounds cannot be directly assimilated by microbial and plant cells, except in some cases such as archaea and bacteria in deep-sea sediments that can recycle membrane lipids and phosphonates from lysed cells (Schouten et al. 2013). In order to access P from organic sources, microorganisms and plants produce phosphatases, which are enzymes that catalyze the transfer of a phosphoryl group from a substrate to water resulting in the hydrolysis of phosphate esters (see also section 1.2.4). Phosphatases are produced by archaea, bacteria, fungi, macrofauna and plants, and are responsible for much of the recycling of organic P in soils.

Phosphatases are commonly classified after their pH optima, metal ion requirements, substrate specificities and reaction mechanisms (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology). Nannipieri et al. (2011) include the following phosphatase classes as relevant for soil organic P hydrolyzis: phosphomonoesterases (EC 3.1.3), phosphodiesterases (EC 3.1.4), triphosphoric monoester hydrolases (EC 3.1.5) and enzymes acting on phosphoryl-containing anhydrides (EC 3.6.1), on P–N bonds (EC 3.9) and on P-C bonds (EC 3.11) (Table 1.1). Phosphomonoesterases can be further classified into acid and alkaline phosphomonoesterase, phosphoprotein phosphatases, phytases (EC 3.1.3.26 for 4-phytase and EC 3.1.3.8 for 3-phytase) and nucleotidases (Table 1.1). Phosphodiesterases include phosphodiesterases

I (EC 3.1.4.1) and phospholipases (EC 3.1.4.3 for phospholipase C and EC 3.1.4.4 for phospholipase D).

While enzymes can be classified according to their biochemical characteristics, enzyme-encoding genes (a pre-requisite of any enzyme) are classified after their amino and nucleic acid sequences. Several classification methods can be used to identify phosphatase-encoding gene families; the Clusters of Orthologous Groups (COG) which is a phylogenetic classification of orthologous genes (Galperin et al. 2014), the Conserved Domain (CD) which uses a reference alignment of conserved amino acids to identify enzyme-encoding gene families (Marchler-Bauer et al. 2011), and the TIGR (The Institute for Genomic Research) gene index (TGI) which uses reference sequences of enzyme-encoding genes with known expressed sequence tags (EST) corresponding to the active residue of the enzyme to identify enzyme-encoding gene families (Quackenbush et al. 2000). According to these classifications, each phosphatase class contains numerous genes that codes for enzymes with different amino acid sequences and co-factors.

1.2.2 Alkaline and acid phosphatases

Phosphomonoesterases such as acid and alkaline phosphomonoesterases are believed to be the most important enzyme class for soil organic P hydrolysis, as soil organic P is mainly composed of phosphomonoesters (Condron et al. 2005). Alkaline phosphatases are mainly produced by prokaryotes, while acid phosphatases are produced by prokaryotes and eukaryotes. In a comparative study based on enzymatic assays, microbial phosphatases have been reported to be more efficient than plant phosphatases (Tarafdar et al. 2002). To date, three alkaline phosphatase genes have been described: *phoA*, *phoD* and *phoX*, coding for the enzymes PhoA, PhoD and PhoX, respectively. Acid phosphatase genes have mainly been studied in single organisms; however, few acid phosphatase gene families such as *acpA* and *appA* have been characterized. Our knowledge about these alkaline and acid phosphatase gene families is summarized below.

Table 1.1. Main phosphatase classes and phosphatase-encoding genes involved in organic P mineralization in soil.

Enzyme class (EC)	Enzyme subclass	Example of genes	Reference	Description	Organisms
Phosphomonoesterases (EC 3.1.3)	Acid phosphatases	*acpA*	TIGR03397	Acid phosphatase, *Burkholderia*-type	Bacteria
		appA	CD08514	Periplasmic substrate-binding acid phosphatase	Bacteria
		AphA	COG3700	Class B non-specific acid phosphatase	Bacteria
		YuiD	COG1963	Membrane-bound protein acid phosphatase	Bacteria, Eukaryotes
	Alkaline phosphatases	*phoA*	COG1785	Alkaline phosphatase	Archaea, Bacteria, Eukaryotes
		phoD	COG3540	Alkaline phosphatase	Archaea, Bacteria, Fungi
		phoX	COG3211	Predicted secreted alkaline phosphatase	Archaea, Bacteria
	Phosphoprotein phosphatases	*Bsu1*	cd07421	*Rhizobiales/ Rhodobacterales*-like phosphatases	Bacteria
		PP1	cd07421	*Rhizobiales/ Rhodobacterales*-like phosphatases	Bacteria
		ApA4	cd07421	*Rhizobiales/ Rhodobacterales*-like phosphatases	Bacteria
	Phytases	*bpp*	pfam02333	Six-bladed beta sheet propeller phytase	Archaea, Bacteria, Fungi
		Phy	COG4247	Myo-inositol-hexakisphosphate 3-phosphohydrolase	Archaea, Bacteria, Eukaryotes
		ptp	pfam14566	PTPlike-phytase	Archaea, Bacteria, Eukaryotes
	Nucleotidases	*SerB*	TIGR00338	N-terminal of phosphoserine phosphatase	Archaea, Bacteria, Eukaryotes
		PG	TIGR01449	2-phosphoglycolate phosphatase	Archaea, Bacteria, Fungi
Phosphodiesterases (EC 3.1.4)	Phosphodiesterase I	*GDPD1*	cd08562	Glycerophosphodiesterase	Bacteria
	Phospholipase D	*PLDP1*	cd08556	Phospholipase	Archaea, Bacteria, Eukaryotes
Triphosphoric monoester hydrolases (EC 3.1.5)		*hflX*	TIGR03156	Probable GTP-binding phosphatase	Archaea, Bacteria, Eukaryotes
Enzymes acting on phosphoryl-containing anhydrides (EC 3.6.1)	Pyrophosphatase	*PPX1*	COG1227	Inorganic pyrophosphatase/exopolyphosphatase	Archaea, Bacteria, Eukaryotes
	Polyphosphatase	*GppA*	COG0248	Exopolyphosphatase/pppGpp-phosphohydrolase	Archaea, Bacteria, Fungi
Enzymes acting on P–N bonds (EC 3.9)	Protein arginine phosphatases	*TPP1*	TIGR01664	Polynucleotide 3'-phosphatase	Archaea, Bacteria, Eukaryotes
		ptpA	pfam04387	Low molecular weight protein-tyrosine-phosphatase	Bacteria, Eukaryotes
Enzymes acting on C–P bonds (EC 3.11)	Phosphonatases	*phnX*	TIGR01422	Phosphonoacetaldehyde hydrolase	Bacteria
	Phosphonopyruvate hydrolase	*Pphn*	TIGR02321	Closely related to phosphoenolpyruvate phosphomutases	Bacteria

PhoA is a homodimeric phosphatase commonly activated by one Mg^{2+} and two Zn^{2+} (Wang et al. 2005; Luo et al. 2010). In some microorganisms such as *Thermotoga maritima* and *Bacillus subtilis*, Co^{2+}, Mn^{2+}, Cd^{2+} can substitute the Zn^{2+} cofactors (Galperin and Jedrzejas 2001; Wang et al. 2005). PhoA acts primarily as phosphomonoesterase (Moura et al. 2001) and reaches its optimal hydrolyzing activity between pH 7.5 and 9.5, depending on the species (Galperin and Jedrzejas 2001; Moura et al. 2001). PhoA alkaline phosphatases are secreted via the Secretory (Sec) pathway (Luo et al. 2010). The *phoA* gene has been found in prokaryotes (archaea, bacteria) and eukaryotes (fungi, insects, mammals) (Millan 1986; Haas et al. 1992; Yang et al. 2000). In bacteria, *phoA* is spread across 26 phyla and is found in high frequency in *Proteobacteria* and *Firmicutes* (Luo et al. 2009). It has been found in many ecosystems such as air, aquatic and terrestrial ecosystems, and also in the microbiome associated to hosts such as annelids, arthropods, mammals and plants (based on the IMG/M database accessed on October 20th 2015). In marine bacteria, *phoA* has been found in *Actinobacteria, Alteromonadales, Bacteroidetes, Burkholderiales, Chlorobi, Gamma-Proteobacteria* and *Rhodobacteriales* (Luo et al. 2009; Sebastián and Ammerman 2009). In fungi, PhoA has been isolated from the saprophytic fungus *Saccharomyces cerevisiae* and the arbuscular mycorrhizal fungus *Gigaspora margarita* (Millan 1986; Narisawa et al. 2003; Yang et al. 2000; Song 2006).

PhoD is a monomeric enzyme that requires one Fe^{2+} and two Ca^{2+} ions as co-factors (Rodriguez et al. 2014). In certain microorganisms, the Fe^{2+} cofactor can be substituted by Mg^{2+} or Zn^{2+} (Gomez and Ingram 1995). The optimal pH of PhoD ranges between 7.5 and 8.5 (Ansai et al. 1998; Yang and Roberts 2004). PhoD is a phosphomonoesterase and also a phosphodiesterase (Kageyama et al. 2011). For example, in the marine bacteria *Aphanothece halophytica*, the primary activity of PhoD is as a phosphomonoesterase with a significant phosphodiesterase activity, while in the gram-negative bacteria *Zymomonas mobilis*, it is the opposite (Yamane and Maruo 1978; Gomez and Ingram 1995). In bacteria, PhoD is exported via the Twin Arginine Translocation (TAT) pathway (Gomez and Ingram 1995). The *phoD* gene has been found in prokaryotes (archaea, bacteria) and

eukaryotes (fungi) (Toh-e et al. 1973; Goldman et al. 1990; Kageyama et al. 2011). In bacteria, *phoD* is mainly found in *Actinobacteria, Proteobacteria* and *Cyanobacteria* (Tan et al. 2013). Moreover, multiple copies of *phoD* can be found in the same genome (Su et al. 2007).

PhoX is a monomeric alkaline phosphatase commonly activated by two Fe^{2+} and three Ca^{2+} as co-factors (Wu et al. 2007; Yong et al. 2014). In some organisms, Fe^{2+}, Mg^{2+} and Cu^{2+} can also serve as co-factors, resulting, however, in a lower activity (Van Mourik et al. 2008; Wu et al. 2007). PhoX reaches an optimum enzymatic activity between pH 7.5 and 10, depending on the organism (Van Mourik et al. 2008; Zaheer et al. 2009; Kathuria and Martiny 2011). For example, the optimum pH of PhoX in *Campylobacter jejuni* is pH 10 (Van Mourik et al. 2008), while it is 7.5 in *Synechococcus* sp. WH8102 (Kathuria and Martiny 2011). The primary activity of PhoX is the hydrolysis of phosphomonoesters. However, the substrate spectrum of PhoX can be larger. The PhoX isolated from *Campylobacter jejuni* has been shown to hydrolyze exclusively phosphomonoesters, whereas the PhoX found in *Sinorhizobium meliloti* can hydrolyze any C-O-P and N-P bonds (Sebastián and Ammerman 2009; Zaheer et al. 2009) and the PhoX in *Pseudomonas multocida* X-73 can hydrolyze both phosphomono- and phosphodiesters. PhoX is secreted via the TAT secretion pathway (Wu et al. 2007; Kathuria and Martiny 2011). The *phoX* gene can be found as part of the genome or on a plasmid as observed in *Rosebacter denitrificans* (Sebastián and Ammerman 2009). To date, the *phoX* gene has only been reported to occur in marine bacteria including *Actinobacteria, Alteromonadales, Bacteroidetes, Cyanobacteria, Lentisphaerae, Planctomycetes* and *Proteobacteria* (Luo et al. 2009; Sebastián and Ammerman 2009). It has also been found in some soil microorganisms such as *Campylobacter jejuni* (van Mourik et al. 2008) and *Sinorhizobium meliloti* (Zaheer et al. 2009).

AcpA is a dimeric acid phosphatase that requires a metal ion as co-factor, although the latter has not yet been identified (Felts et al. 2006). The protein structure resembles that of PhoD and PhoX alkaline phosphatase (Felts et al. 2006). AcpA is defined as a nonspecific

acid phosphatase and has an optimal activity at pH 6 to 7 (Costas et al. 2010; Reilly et al. 1996). There is evidence that AcpA can hydrolyze pyrophosphate, glycerophosphodiesters, phospholipids, single nucleotides (e.g. AMP, ATP) and phosphomonoesters (e.g. Glucose-, fructose- or ribose-6 phosphate) (Felts et al. 2006). AcpA is exported to the periplasm by the TAT pathway (De Buck et al. 2008; Felts et al. 2006). The *acpA* gene has first been isolated and characterized from pathogenic bacterial strains such as *Francisella tularensis* (Reilly et al. 1996). Analogue genes of *acpA* have been found in other bacterial genomes affiliated to *Proteobacteria (Alpha- and Gamma-Proteobacteria), Actinobacteria, Bacteroidetes, Chlamydiae, Cyanobacteria, Firmicutes* and *Spirochaetes* (Costas et al. 2010). Our knowledge of the prevalence of *acpA* in nature, however, remains poor.

AppA1 and AppA2 are monomeric acid phosphatases with an optimum pH of 2.5 and 4.5, respectively, and are stable between pH 2 and 10 (Dassa and Boquet 1985; Golovan et al. 1999; Mullaney and Ullah 2003). Both AppA are primarily phosphomonoesterases but exhibit also phytase activity in certain species (Golovan et al. 1999; Mullaney and Ullah 2003). They are mainly found in the periplasm via the Sec pathway (Dassa et al. 1990; Matos et al. 2014). The *appA* genes have first been characterized in *Escherichia coli* and have been found by analogy in other bacteria affiliated to *Chlamydiae, Chlorobi, Deferribacteres, Firmicutes, Proteobacteria (Alpha-, Beta-, Gamma-, Delta-* and *Zeta-Proteobacteria)* and *Planctomycetes* (NCBI, Conserved Domain Database search). Additionally, AppA1 has been found in *Schizosaccharamyces pombe, Saccharomyces cerevisiae, Aspergilus niger*, rat and human cells, although it has been given different names (Piddington et al. 1993).

1.2.3 Phosphatase gene regulation

To respond rapidly and appropriately to variations in P availability, while minimizing the costs and resources for the cell, phosphatase synthesis is subject to fast and sensitive

genetic regulation (Vershinina and Znamenskaya 2002). In bacteria, phosphatases are commonly part of the Pho regulon, a general regulatory system for P management in the cell that includes a collection of phosphate-starvation inducible genes, such as genes encoding for phosphatases, orthophosphate-specific transporters, enzymes involved in P storage and other P mobilizing systems (Vershinina and Znamenskaya 2002). These genes can be organized in operons, which are clusters of genes under the control of a single promotor, or in single genes with their own promotor. The regulation of the expression of the *pst* genes occurs via a two-component regulatory system including a transmembrane sensory histidine protein kinase and a regulatory protein that interacts with a specific part of the promotor sequence called the Pho BOX (Santos-Beneit 2015). In bacteria, the two-component signal transduction system is often composed of PhoB-PhoR molecules in gram-negative bacteria and of PhoP–PhoR in gram-positive bacteria (Vershinina and Znamenskaya 2002). However, the promoter sequence and the regulatory proteins can vary greatly from one species to another and have not yet been fully elucidated (Santos-Beneit 2015). For example, in *Escherichia coli*, the Pho regulon regulates 137 genes, 118 of which are induced and the other 19 are suppressed by phosphate starvation conditions (Van Bogelen et al. 1996). The *phoA*, *phoD* and *phoX* alkaline phosphatase genes and the *acpA* and *appA* acid phosphatase genes are all induced under phosphate starvation conditions. However, the regulation of the *phoA*, *phoD* and *phoX* alkaline phosphatase genes and the *acpA* acid phosphatase gene are positively controlled, meaning that a signal molecule binds to the promotor to activate the expression of the gene, while the *appA* acid phosphatase genes are negatively controlled, meaning that a repressor protein prevents a gene from being expressed (Eder et al. 1996; Stonehouse et al. 2002; Touati and Danchin 1987; Van Mourik et al. 2008; Wang et al. 2005). In a P deprivation study on *Synechococcus* sp. OS-A and OS-B isolated from bacterial mats in the Yellowstone National park which harbor all three alkaline phosphatase genes, *phoX* was up-regulated 200-fold within the first 72 hours after P depletion, while *phoA* and *phoD* were up-regulated 6- and 21-fold, respectively (Adams et al. 2008). In their

exponential growth phase, bacteria can up-regulate their phosphate-starvation inducible genes within 4 min, while genes involved in iron-sulfur protein assembly, nucleotide metabolism, lipopolysaccharide biosynthesis, and aerobic respiration can take up to 20 min (Rolfe et al. 2012).

1.2.4 Reaction mechanisms of phosphatases

Alkaline and acid phosphatases catalyze principally the hydrolysis of phosphomonoesters (Figure 1.2a). They all have in common the presence of a positive charge at the active sites that enables the attack of the highly negatively charged nucleophile that is composed of the four oxygen atoms bound to P (Hengge 2005). The reaction mechanism differs between alkaline and acid phosphatases. While all known alkaline phosphatases have conserved metal ion centers, which differ between PhoA, PhoD and phoX, the acid phosphatases generally do not require co-factors for the catalysis (Hengge 2005).

For example, the PhoA alkaline phosphatase requires two Zn^{2+} and one Mg^{2+} and proceeds via an intermediate in which a serine residue is phosphorylated (Von Sperber et al. 2014). In more detail, the first step of the reaction is the stabilization of the phosphomonoester by the enzyme, which is coordinated by the two Zn^{2+} co-factors (Figure 1.2b). Once the substrate is bound, the active serine residue acts as a nucleophile and attacks the phosphoryl group of the phosphomonoester, resulting in the serine phosphorylated intermediate and the release of the alcohol moiety. Coupled with the release of the organic molecule, a hydroxide group from the surrounding water molecules takes the coordination site of one of the Zn^{2+}, displacing the phosphoryl group from the active residue and releasing the orthophosphate. Although the role of Mg^{2+} in the reaction is not clear, the Mg^{2+} is believed to function as the provider of a general base that deprotonates the active serine residue. As a result, three O atoms are inherited from the original phosphomonoester substrate, while the fourth O atom is incorporated from

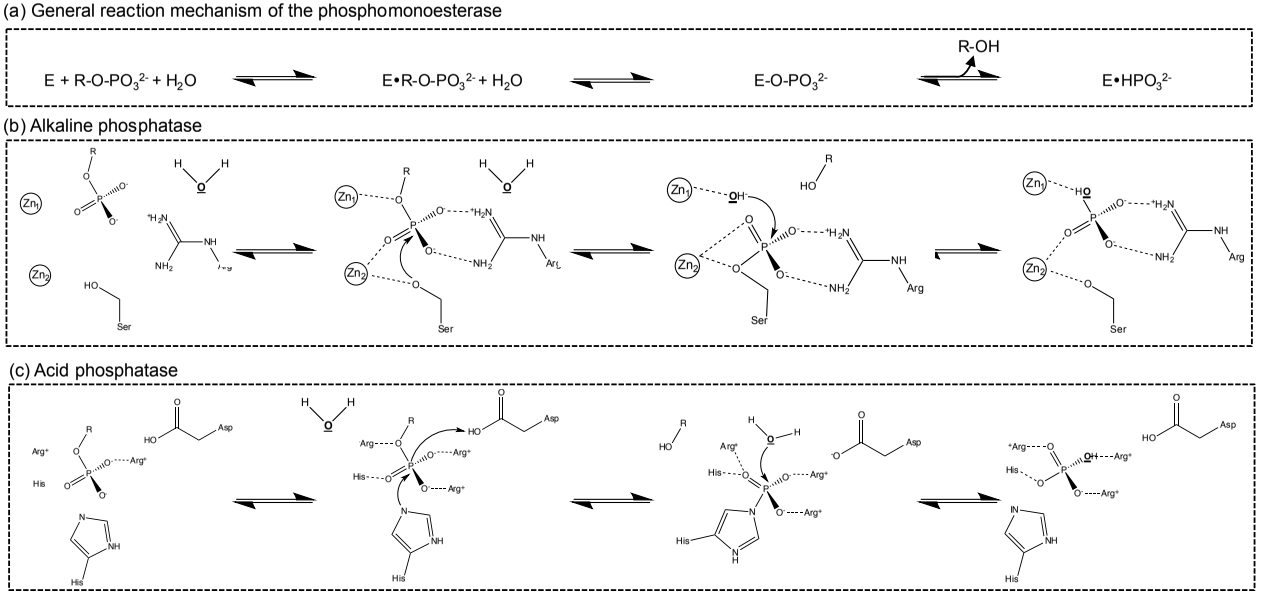


Figure 1.2. (a) Principle of reaction mechanism of phosphomonoesterases such as the alkaline and acid phosphatase (adapted from Hengge (2005)). (b) Reaction mechanism at the active site of alkaline phosphatases. The substrate is stabilized and then bound to the active site of the enzyme. A serine residue attacks the phosphoryl group resulting in the formation of a covalent enzyme–phosphate intermediate and the release of the leaving group. Subsequently, a hydroxide ion from the surrounding water takes the coordination site at Zn1 and displaces the phosphorus group from the serine, forming the enzyme–phosphate complex. (c) Reaction mechanism at the active site of acid phosphatase. Once the substrate is bound to the enzyme, a nucleophilic attack of the histidine on the phosphorus leads to the formation of a covalent enzyme–phosphate intermediate and release of the leaving group. Subsequently, the carboxylate of the aspartate deprotonates a water molecule, creating a nucleophilic hydroxide ion, which displaces the phosphoryl group from the histidine forming an enzyme–phosphate complex.

the hydroxide during the hydrolysis of the phosphorylated intermediate (Liang and Blake 2006).

In contrast to the PhoA alkaline phosphatases, acid phosphatases generally require positively charged amino acids for the catalysis and do not incorporate a metal-coordinated hydroxide ion, but a hydroxide directly derived from a water molecule (Figure 1.2c). In the case of an acid phosphatase isolated from mammals, likely corresponding to AppA acid phosphatase (Piddington et al. 1993), three arginines, one histidine and one aspartate catalyze the hydrolysis of phosphomonoesters (Liang and Blake 2006; Von Sperber et al. 2014). The three arginines stabilize and orient the phosphoryl group of the phosphomonoester to allow the nucleophile histidine to selectively bond with the phosphoryl group. This results in the formation of the phosphorylated intermediate and the release of the organic molecule. In concert with the nucleophilic substitution, the aspartate deprotonates a water molecule, which consequently displaces the phosphoryl group from the phosphorylated histidine intermediate, releasing the orthophosphate into the soil solution.

1.2.5 Location of phosphatases in microorganisms and in soil

Phosphatases can be intracellular and extracellular. Intracellular phosphatases are found in the cytoplasm, where they recycle the organic P within the cell. Extracellular phosphatases are secreted via the TAT pathway dedicated to the transport of proteins in their folded state or the Sec-pathway translocating proteins in their unfolded state. They can be found in the periplasm or bound to the outer-membrane, or can be fully secreted into the environment (Luo et al. 2009), which corresponds to different ecological strategies. Phosphatases can be inactivated or inhibited by physicochemical soil properties (e.g. pH, moisture), adsorbed and degraded. If the optimal soil conditions are met, there is still a low chance that they will encounter the right substrate which itself is often sorbed on organomineral surfaces. As a result, phosphatase activity in soil refers to a composite

of activities associated with various biotic and abiotic soil components. These can be categorized into the 11 categories (Burns 1982; Burns et al. 2013) illustrated in Figure 1.3. The location of enzymes in the soil is determined by multiple factors such as the enzyme size and the solubility in the soil solution, and changes over time. Some phosphatases such as the membrane-bound and extracellular phosphatases can persist at least a month (Pettit et al. 1977; George et al. 2005) and up to several months to years as some studies suggest (Li et al. 1998; Quiquampoix and Mousain 2005), while others such as the cytoplasmic phosphatases are very unstable once released into the environment through cell lysis (Burns et al. 2013). Together with the variety of organic P forms, this makes the interpretation of phosphatase activity measurements difficult.

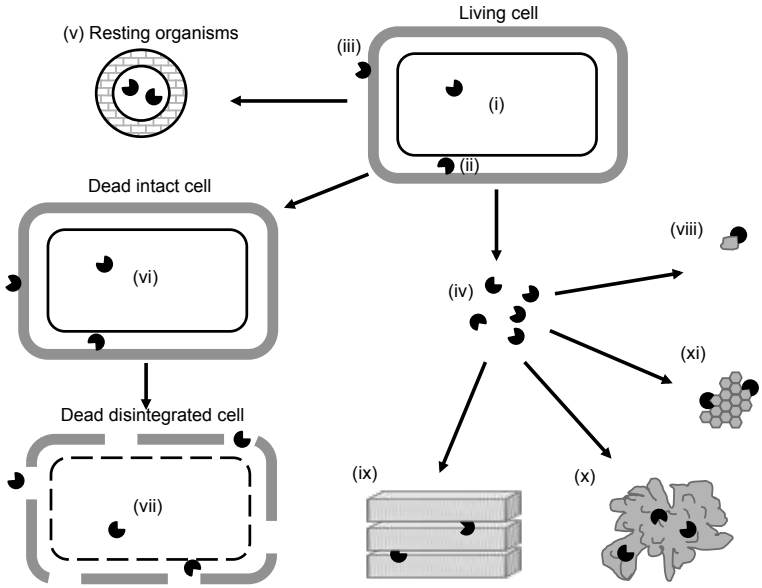


Figure 1.3. Locations of enzymes in soil (adapted from Burns et al. (2013)): (i) in the cytoplasm of microbial cells; (ii) in the periplasm of gram-negative bacteria; (iii) bound to the outer membrane; (iv) free in the soil aqueous phase; (v) in resting/dormant cells; (vi) attached to entire dead cells and cell debris; (vii) leaking from intact cells or released from lysed cells; (viii) associated with enzyme-substrate complexes; (ix) sorbed to clay mineral surfaces; (x) complexed with soil organic matter by adsorption, or entrapment; (xi) bound to condensed tannins.

1.3 How to study phosphatases in soil

1.3.1 Phosphatase activity

Phosphatases were first studied by measuring their potential activity in assays which consist of the addition of *p*-nitrophenyl phosphate as substrate in excess to a soil suspension buffered at the desired pH that turns yellow at higher pH after catalysis and of the measurement of the increase in absorbance overtime (Burns 1978; Weaver et al. 1994). This colorimetric approach is less sensitive than the more recent use of fluorescent substrates (Marx et al. 2001). The potential acid and alkaline phosphatase activities are generally measured at pH 6.1 and pH 11, respectively, while the effective phosphatase activity is measured at soil pH to mimic natural conditions (Marx et al. 2001; Poll et al. 2006). This method has, however, drawbacks, which are (i) the inevitable soil disturbance that will modify the microbial activity and thus phosphatase activity, (ii) the substrate that is artificial and represents one type of substrate only, the fact that (iii) the type of buffer and (iv) the pH will both strongly affect the enzyme activity (Burns et al. 2013). Moreover, the in vitro conditions cannot reflect the heterogeneous in situ conditions, and thus the interpretation of potential phosphatase activity measurements can be problematic. New methods are emerging for the measurement of enzymatic activity such as microarray biochips which detect enzymes by using high-affinity antibodies and a subsequent immunofluorescent assay (Rivas et al. 2008). Fluorescent-Activated Cell Sorting (FACS) measures fluorescence of single cells by flow cytometry and could be used to sort phosphatase-producing cells by using a fluorescent substrate, such as ELF® 97 phosphate which precipitates after hydrolysis by a phosphatase at the site of enzymatic activity (Paragas et al. 2002). These promising methods have, however, not yet been adapted to phosphatases in environmental samples, especially in soil.

1.3.2 Isotopic tracers

Isotopic tracers can be used as tools to track P during transfers between soil pools and to determine the rates of abiotic and biotic processes such as gross and net organic P mineralization. Phosphorus has one stable (^{31}P) and two radioactive isotopes (^{32}P, ^{33}P) that are commonly used to study P cycling. Oehl et al. (2001) proposed a method to measure gross organic P mineralization based on isotopic dilution principles. In the isotopic dilution approach, a tracer (^{32}P or ^{33}) is added to a soil-solution system to follow the dynamics of ^{31}P (the tracee) (Fardeau 1993). In a system labelled with ^{32}P or ^{33}P, the release of inorganic ^{31}P from organic ^{31}P decreases the specific activity of P (^{32}P/^{31}P or ^{33}P/^{31}P) in the soil solution (Bünemann 2015). To determine gross organic P mineralization, first, the decrease in the specific activity of the soil solution after addition of the P isotope has to be measured in a short-term batch experiment (Bünemann et al. 2007), during which the specific activity of phosphate in the soil solution can be assumed to result solely from physicochemical processes, and can then be extrapolated for longer times. Secondly, gross organic P mineralization rates are determined by comparison of isotopically exchangeable P (E-values) extrapolated from the short-term batch experiment to those measured in an incubation experiment of 7 to 14 days duration using incubated non-sterile soil (Bünemann et al. 2007; Bünemann 2015), in which isotopic dilution results from physicochemical as well as biological processes. Finally, microbial P immobilisation is derived from microbial tracer uptake measured in the incubation experiment and net organic mineralization rates are calculated as the difference of gross mineralization and microbial immobilisation rates (Bünemann et al. 2007).

The oxygen (O) isotope composition in phosphate (δ^{18}O-PO_4) has been proposed as a new tool to study biological processes in the P cycle (Tamburini et al. 2010). In contrast to the radioactive tracers ^{32}P and^{33}P, δ^{18}O-PO_4 can provide insight into the mechanisms and transformations during an enzymatic reaction. The advantages of the use of δ^{18}O-PO_4 are

that (i) O exchange between phosphate and water is slow and negligible in abiotic systems (Blake et al. 2005), thus, O will preserve its original isotopic composition in the absence of biological activity, (ii) processes such as adsorption–desorption and precipitation in soil do not produce any significant O isotope fractionation (Liang and Blake 2007; Jaisi et al. 2010), (iii) microorganisms preferentially take up lighter isotopologues of phosphate, leading to an enrichment of heavier isotopologues in the residual phosphate (Blake et al. 2005), and (iv) the hydrolysis of phosphoesters by phosphatases will release phosphate with a different isotopic composition than the original organic compound (Liang and Blake 2006). Acid phosphatases from wheat germ and potato and alkaline phosphatase from *Escherischia coli* have been reported to have a distinct O isotope fractionation (Von Sperber et al. 2014). The O isotope fractionation caused by acid phosphatases is 20–30 ‰smaller than for alkaline phosphatases (Von Sperber et al. 2014). As a result, the O isotope composition of inorganic P released during the enzymatic reaction can be used to trace different enzymatic processes.

1.3.3 Proteomics

Phosphatases can be studied at the proteomic level that consists of the analysis of all proteins in a sample, including free, sorbed and entrapped enzymes, using reversed-phase liquid chromatography coupled to mass spectrometry (LC–MS). The main challenges of this technique are the improvement of the extraction efficiency of proteins from environmental samples, the accurate identification of the protein for non-model organisms and the development of data analysis tools that handle the heterogeneity and complexity of microbial communities in the environment (Dowd 2012).

1.3.4 Genomics and transcriptomics

Phosphatases can be studied at the genetic level, representing the potential production of phosphatases, and at the transcriptomic level, representing the expression of phosphatase genes via two main approaches; shotgun metagenomics and gene-targeted metagenomics. Shotgun metagenomics consist of the sequencing of all DNA present in a given environmental sample including DNA from virus particles, prokaryotic and eukaryotic cells and free DNA. This approach has the great advantage to amplify DNA non specifically, giving the chance to discover novel species and functional genes. However, at DNA level, housekeeping genes such as ribosomal genes are much more abundant than functional genes, diluting the functional genes in the DNA pool and making their detection less likely. The second approach is a gene-targeted technique, which uses primers. The use of primers allows amplification of a gene of interest with high sensitivity. Hence, even at low abundance of the gene in the sample, the amplification results in a more comprehensive and complete inventory of the gene of interest. These two approaches can be used to study phosphatase-encoding transcripts at the RNA level (metatranscriptomics). However, the success of the study depends entirely on the design of the primers. The gene-targeted approach represents currently the best technique in terms of resolution, sensitivity and reliability to study functional genes in environmental samples.

1.4 Genetic tools targeting phosphatase genes

1.4.1 Primer design

Primer design consists of the selection of a pair of oligonucleotides used to target conserved regions in the gene of interest. The aim in primer design is to obtain a balance between specificity, coverage and efficiency of amplification. Specificity is defined as the

frequency with which a mispriming event occurs (Dieffenbach et al. 1993). Coverage describes how well the primers amplify the gene of interest given the diversity of the genes in the targeted population. For many gene families, the nucleic and amino acid sequence is not well conserved, resulting in a large variety that can only be addressed by either the use of multiple primers or degenerate primers. Finally, efficiency is defined as the measured exponential increase in amplicon per cycle compared to the theoretical optimal exponential increase in amplicon per cycle.

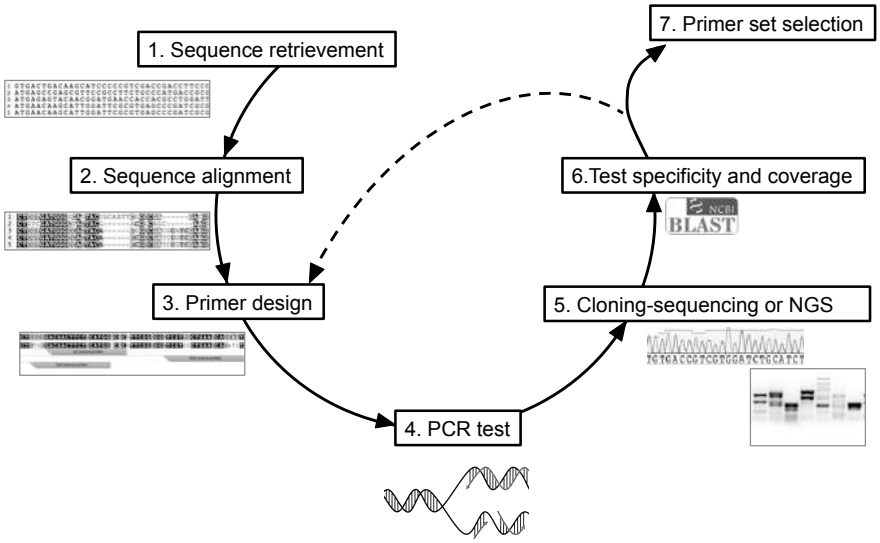


Figure 1.4. Steps in primer design. (1) Sequences are retrieved and (2) aligned. (3) Primers are designed and (4) tested in PCR. (5) PCR products are sequenced. (6) Specificity and coverage of the primers are evaluated. If the coverage of specificity is poor, new primers are designed (3). If the coverage and specificity are satisfying, the primer sets are selected for further studies (7).

The steps in primer design are shown in Figure 1.4: (1) All available sequences of the gene of interest are retrieved from databases such as NCBI, ENA, UniProtKB and Mega-RAST. (2) The sequences are aligned using an alignment program such as MUSCLE, MUltiple Sequence Comparison by Log-Expectation (Edgar 2004), and

revised manually to reach the optimal alignment. (3) Primers are designed in two steps; first by computer analysis that locates conserved regions in the alignment and designs primers with appropriate melting temperatures. These primers are then manually improved to reach the best coverage. (4) Primers are paired depending on their fragment size and annealing temperature, and tested in a Polymerase Chain Reaction (PCR) on environmental samples and strains. (5) The PCR products of successful amplification are further tested for specificity and coverage by cloning-Sanger sequencing or Next Generation Sequencing (NGS). (6) Evaluation of the specificity and coverage is done using public databases like NCBI. If the coverage and specificity are poor, new primers are designed and tested until appropriate coverage and specificity are reached. If the coverage and specificity are satisfying, the primer sets are selected for applied studies (7).

1.4.2 Phosphatase gene-targeting primers

Only few primers have been designed that target phosphatase-encoding genes and are suitable for environmental studies. Sakurai et al. (2008) designed the first primers that target the *phoD* gene in soil microorganisms. Using these primers, the *phoD*-harboring community has been studied in many soils and has been reported to be affected by fertilization type and intensity, vegetation and certain soil properties such as organic carbon content and pH (Wang et al. 2012b; Chhabra et al. 2013; Jorquera et al. 2014; Cui et al. 2015). Additionally, a positive correlation between *phoD* gene abundance and potential alkaline phosphatase activity was observed in a long-term trial including non-fertilized and fertilized soils (Fraser et al. 2015b), suggesting that PhoD contributes significantly to the total alkaline phosphatase activity. Nonetheless, these primers have been shown to have an amplification bias, resulting in an overrepresentation of *Alpha-Proteobacteria* (Tan et al. 2013). New primers are therefore required to provide a better coverage of the *phoD* diversity. The first primers targeting the *phoX* gene were designed by Sebastián and Ammerman (2009) based on marine bacterial sequences such as *Vibrio*

sp., *Shewanella* sp. and *Roseobacter* sp. and were used to assess the abundance of *phoX* in the Sargasso Sea and the Chesapeake Bay. Using the same primers, Dai et al. (2014) showed that *phoX* was common in mesotrophic and hypereutrophic waters of Lake Taihu (China) and that different P concentrations select for different *phoX*-harboring microorganisms. No primers that target other alkaline and acid phosphatase genes in soil microorganisms have yet been developed. In a metagenomic study, Bergkemper et al. (2015) reported that the community harboring phosphatase genes can differ between a P-limited and a P-rich forest soil. They observed that *Rhizobiales* were dominant in the P-rich soil, while *Actinomycetales*, *Acidobacteriales* and *Solibacterales* were dominant in the P-limited forest soil. They also showed that a large fraction of phosphatase genes are found in poorly described bacterial and fungal orders.

On the whole, knowledge on the taxonomic distribution and prevalence of phosphatase-encoding genes in the soil environment is missing and the development of new molecular tools could be a solution to address this issue.

1.5 Outline and objectives

The aims of this thesis are to develop and apply new molecular tools targeting phosphatase gene families to soils. New sets of primers are designed and validated for *phoD* and *phoX* alkaline phosphatase genes (Chapters 2 and 3). Using these newly designed primers, the key microorganisms that harbor the *phoD* and *phoX* genes are identified, and the prevalence and relationship between environmental factors and the *phoD* and *phoX* gene in 30 soils across different land-uses, climates and soil types in Australia and Switzerland are explored (Chapter 4). Finally, the effects of phosphate depletion and soil pH on both the total and active *phoD*-harboring microorganisms are investigated in a long-term grassland fertilization trial (Chapter 5).

The general hypotheses of this thesis are (i) that the taxonomic distribution of *phoD* and *phoX* is on the whole similar as some studies showed that *phoD* and *phoX* are

often found in the same genomes (Zaheer et al. 2009; Sebastián and Ammerman 2011), (ii) that dominant *phoD*- and *phoX*-harboring microorganisms vary depending on the environmental conditions and more particularly depending on P availability as low P availability may promote specific *phoD*- and *phoX*-harboring microorganisms and on pH as both PhoD and PhoX reach their optimum hydrolyzing activity at alkaline pH, and finally (iii) that the active *phoD*-harboring community represents only a limited number of taxa of the total *phoD*-harboring community and is more responsive to phosphate depletion and differences in soil pH than the total *phoD*-harboring community as it has been reported that only some members of the total community are metabolically active during decomposing processes (Baldrian et al. 2012).

2

PHOD ALKALINE PHOSPHATASE GENE DIVERSITY IN SOIL

Published as:

Ragot SA, MA Kertesz and EK Bünemann (2015). *phoD* alkaline phosphatase gene diversity in soil. *Applied and Environmental Microbiology* 81, pp. 7281-7289.

Abstract

Phosphatase enzymes are responsible for much of the recycling of organic phosphorus in soils. The PhoD alkaline phosphatase takes part in this process by hydrolyzing a range of organic phosphoesters. We analyzed the taxonomic and environmental distribution of *phoD* genes using whole genome and metagenome databases. *phoD* alkaline phosphatase was found to be spread across 20 bacterial phyla and was ubiquitous in the environment, with the greatest abundance in soil. To study the large diversity of *phoD*, we developed a new set of primers which targets *phoD* genes in soil.

The primer set was validated on six soils collected from two continents with different climates and soil properties, using 454-sequencing, and compared to previously published primers. Up to 685 different *phoD* operational taxonomic units were found in each soil, which was seven times higher than with previously published primers.

The new primers amplified sequences belonging to 13 phyla including 71 families. The most prevalent *phoD* genes identified in these soils were affiliated with the orders *Actinomycetales* (13-35%), *Bacillales* (1-29%), *Gloeobacteriales* (1-18%), *Rhizobiales* (18-27%) and *Pseudomonadales* (0-22%). The primers also amplified *phoD* genes from additional orders, including *Burkholderiales, Caulobacterales, Deinococcales, Planctomycetales* and *Xanthomonadales*, which represented the major differences in *phoD* composition between samples, highlighting the singularity of each community. Additionally, *phoD* bacterial community structure was strongly related to soil pH, which varied between 4.2 and 6.8.

These primers reveal the diversity of *phoD* in soil and represent a valuable tool for the study of *phoD* alkaline phosphatase in environmental samples.

2.1 Introduction

Phosphorus (P) is an essential macronutrient for all living cells (Westheimer 1987). Despite its relative abundance in soils, P is one of the main limiting nutrients for terrestrial organisms (Vitousek et al. 2010). Phosphorus is present in organic and inorganic forms in soil, but only the inorganic orthophosphate ions in soil solution are available for plants (Vitousek et al. 2010). To sustain crop productivity, P fertilizers are therefore used in agriculture, both as inorganic fertilizers (e.g. triple super phosphate) and organic fertilizers (e.g. manure). After application, some of the inorganic P is rapidly taken up by plants and microorganisms, while the remaining P is immobilized as insoluble and bound P forms in the soil. Microorganisms can access and recycle P from these recalcitrant P forms by solubilization of inorganic P and by mineralization of organic P via enzymatic processes mediated primarily by phosphatases, which hydrolyze the orthophosphate group from organic compounds (Vitousek et al. 2010). When facing P scarcity, microorganisms up-regulate expression of functional genes coding for phosphatases (phosphomonoesterases, phosphodiesterases, phytases), high affinity phosphate transporters, and enzymes for phosphonate utilization, which together constitute the Pho regulon (Vershinina and Znamenskaya 2002). The phosphomonoesters which are hydrolyzed by phosphatases are generally the dominant fraction of organic P and can represent up to 90% of the organic P in soil (Vitousek et al. 2010).

Prokaryotic alkaline phosphatases have been grouped into three distinct families, PhoA, PhoD and PhoX (Boulanger and Kantrowitz 2003; Wu et al. 2007; Kageyama et al. 2011), which are classified in COG1785, COG3540 and COG3211, respectively, of the Cluster of Orthologous Groups (COG) categorization. PhoA was the first alkaline phosphatase to be characterized. It is a homodimeric enzyme that hydrolyzes phosphomonoesters, and is activated by Mg^{2+} and Zn^{2+} (Boulanger and Kantrowitz 2003). PhoD and PhoX are monomeric enzymes that hydrolyze both phosphomonoesters and phosphodiesters and are activated by Ca^{2+} (Wu et al. 2007; Kageyama et al. 2011). Enzymes of all three

families are predominantly periplasmic, membrane-bound or extracellular (Luo et al. 2009). PhoD and PhoX are exported by the twin-arginine translocation pathway (Wu et al. 2007; Kageyama et al. 2011), while PhoA is secreted via the Sec protein-translocation pathway (Zaheer et al. 2009). There is high sequence variability in the PhoA, PhoD and PhoX proteins, not only between the families, but also within each family (Zaheer et al. 2009; Kageyama et al. 2011). PhoD is widespread in both terrestrial and aquatic ecosystems (Luo et al. 2009; Tan et al. 2013).

Until recently, our knowledge of the phosphatase-encoding genes in prokaryotes was based on traditional culture-dependent methods. Advances in culture-independent techniques have provided new tools for the study of microbial communities in the environment. The first functional gene probes to target alkaline phosphatase genes were the ALPS primers developed by Sakurai et al. (2008). They were based on phosphatase gene sequences from seven isolates and first used to examine the different soil alkaline phosphatase community structures resulting from mineral and organic fertilization. Alkaline phosphatase genes belonging to the *Actinobacteria*, *Alpha-*, *Beta-* and *Gamma-Proteobacteria*, and *Cyanobacteria* were identified by cloning, giving the first insight into alkaline phosphatase diversity in soil (Sakurai et al. 2008).

Subsequently, the ALPS primers were demonstrated to be specific to the *phoD* alkaline phosphatase gene (Tan et al. 2013). They were used to assess alkaline phosphatase gene diversity and structure in several soils by PCR-DGGE (Wang et al. 2012a; Wang et al. 2012b; Chhabra et al. 2013; Jorquera et al. 2014) and by 454-sequencing (Fraser et al. 2015a). These studies showed that crop management, application of organic and conventional fertilizers and vegetation all affect the *phoD* alkaline phosphatase gene diversity. Tan et al. (2013) examined the effect of three mineral P fertilization intensities (zero, medium and high input) in grassland soil on the composition and diversity of alkaline phosphatase, and found a change in the *phoD* bacterial community composition between unfertilized and fertilized treatments, with the dominant *phoD* alkaline phosphatase genes affiliated to *Alpha-* and *Gamma-Proteobacteria*, *Actinobacteria* and

Cyanobacteria. However, they pointed out that the ALPS primers are likely to have an amplification bias resulting in an over-representation of *Alpha-Proteobacteria*, and that new primers are therefore required to provide better coverage of the *phoD* diversity.

In this study, we assessed the diversity and environmental distribution of the *phoD* gene based on current genome and metagenome databases, and we present a new set of improved primers which targets the large diversity of *phoD* genes in soil microorganisms. These primers can be used both as a tool to identify PhoD-producing bacteria and to study *phoD* bacterial community diversity and composition in the environment. The newly designed primers were tested in a gene-targeted metagenomic approach using 454-sequencing in a range of soils collected from two continents with different climates and soil properties. Finally, we compared them to the previously published ALPS primers (Sakurai et al. 2008), using the same samples and methodology.

2.2 Material and Methods

2.2.1 Taxonomic and environmental distribution of *phoD* alkaline phosphatase genes across microbial genomes and metagenomes

The distribution of *phoD* genes was assessed using the Integrated Microbial Genomes and Metagenomes (IMG/M) database, a dedicated system for annotation of whole genomes and metagenomes (Markowitz et al. 2012). Draft and complete genome datasets were used to evaluate the distribution of *phoD* across kingdoms and phyla, and metagenome datasets were used to evaluate the prevalence of *phoD* in the environment (data accessed on July 13th 2015). Metagenome datasets were categorized into “air”, “engineered and waste” (bioreactor and waste treatment), “extreme environments” (saline, alkaline, hot spring, brine and black smokers), “fresh water”, “marine environment”, “plant-associated” (leaves and wood), “animal-associated” (associated to humans, arthropods, molluscs and sponges) and “soil” (rhizosphere and bulk soil). These categories were chosen based on

the environment type classification of the IMG/M database. The relative abundance of *phoD* gene counts per environment type was calculated as gene count number normalized by the total number of bases sequenced per metagenome dataset.

2.2.2 Soil sampling and general soil characteristics

Four grassland soils were collected in Australia in July 2013 (S1-4) and two grassland soils were sampled in Switzerland in September 2012 (S5-6) (Table 2.1). These represent a broad range of soil types, vegetation and climatic conditions varying from hot semi-arid to continental temperate climates. At each site, five soil cores from the top 5 cm were randomly collected and homogenized by sieving (4 mm). A subsample was stored at -80°C for molecular analysis. The remaining composite soil was air-dried and used to determine basic soil properties including pH, texture, total C and P. Methods used to determine the soil properties are described in Table 2.1. The sampled soils covered a range of textures, with clay contents varying between 12 and 38%. Soil pH ranged between 4.2 and 6.8. Total C varied between 5 and 34 g kg^{-1} soil, and total P between 193 and 705 mg kg^{-1} soil. The vegetation was similar at sampling sites S5 and S6, and very different in the other sites, ranging from dense to scarce vegetation, depending on the location.

2.2.3 DNA extraction from soil

All DNA samples were extracted in duplicate. Nucleic acids were extracted from the Australian samples using the DNA Powersoil Isolation kit® (MO BIO, Carlsbad, CA, USA) according to the manufacturer's instructions, with an initial bead beating step of 2 cycles of 3 min at 30 Hz using a TissueLyzer II (Qiagen, CA). Nucleic acids were extracted from the Swiss samples from 2 g of frozen soil using the RNA Powersoil Isolation kit® (MO BIO, Carlsbad, CA, USA) according to the manufacturer's instructions, with

Table 2.1. Description of the grassland soils S1 to S6, with location, geographical coordinates, climate, soil type, vegetation, pH, texture, and total C and P.

Sample	Site	Geographical coordinates	Climate[1]	Soil type[2]	Vegetation	pH[3] ($CaCl_2$)	Texture[4] (clay, silt, sand in %)	Total C[5] g kg^{-1} soil	Total P[6] mg kg^{-1} soil
S1	Kia-Ora (Australia)	34°48'18"S, 148°35'00""E"	Warm temperate climate, fully humid with warm summer (Cfb)	Planosol	*Austrodanthonia* spp., *Austrostipa* spp. *Elymus scaber*, *Microlaena stipoides*, *Bothriochloa macra*,	4.2 ± 0.3	14, 28, 58	21 ±0.8	221± 8
S2	Narrabri (Australia)	30°15'14"S, 149°51'53"E	Warm temperate climate, fully humid with warm summer with hot summer (Cfa)	Planosol	*Chrysocephalum* sp., *Festuca arundinacea* *Themada* sp.,	6.1 ± 0.0	38, 27, 35	23.7 ± 0.1	705 ± 13
S3	Nyngan (Australia)	31°25'52"S, 147°04'09""E"	Arid climate, hot steppe (BSh)	Cambisol	Mixed grasses and dicot plants. Clumpy cover, not a sward.	4.7 ± 0.1	30, 33, 37	15 ± 0.3	466 ± 10
S4	Mutawintji (Australia)	31°16'19"S, 142°17'44""E"	Arid climate, hot steppe (BSh)	Leptosol	*Chenopodium* sp., *Astrebla* sp. *Acacia* sp.,	6.8 ± 0.1	12, 11, 77	5 ± 0.2	193 ± 11
S5	Watt (Switzerland)	47°25'45"N, 008°29'31"E	Warm temperate climate, fully humid with warm summer (Cfb)	Cambisol	*Arrhenaterion elatioris*	5 ± 0.1	30, 33, 37	27.5 ± 0.1	613 ± 33
S6	Watt (Switzerland)	47°25'45"N, 008°29'31"E	Warm temperate climate, fully humid with warm summer (Cfb)	Cambisol	*Arrhenaterion elatioris*	6.1 ± 0.3	30, 33, 37	34.4 ± 0.4	703 ± 39

[1]Climate classification (Kottek et al. 2006).
[2]World Reference Base for Soil Resources (IUSS 2014).
[3]measured in a soil suspension in 0.01 M $CaCl_2$ with a 1 to 2.5 mass to volume ratio using a Benchtop pH/ISE 720A (Orion Research Inc., Jacksonville, FL).
[4]was determined by a commercial soil analysis lab (Soil Conseil, Nyon, Switzerland).
[5]measured on dry and ground soil using a CNS analyzer (Thermo-Finnigan).
[6]determined by wet digestion with H_2O_2/H_2SO_4 (Anderson and Ingram 1993) and measured with malachite green at 610 nm (Ohno and Zibilske 1991).

an additional homogenizing step using the Omni Bead Ruptor Homogenizer (Omni International, Kennesaw, GA) (2.8 mm zirconium beads for 1 min at 5 m s^{-1}) prior to isolation. DNA was eluted from the RNA/DNA Capture Column using 4 ml of DNA elution solution (1 M NaCl, 50 mM MOPS, 15% isopropanol, pH 7). DNA was precipitated using isopropanol, and resuspended in DEPC-treated H_2O. Only the DNA extracts were used in this study.

2.2.4 Primer design and in silico testing

Gene sequences annotated as *phoD* and/or associated with COG3540 (Cluster of Orthologous Groups; http://ww.ncbi.nlm.nih.gov/COG/), which corresponds to *phoD* alkaline phosphatase, were retrieved from the European Nucleotide Archive (ENA) and Uniprot Knowledgebase (UniprotKB) databases. They were then clustered at 97% similarity using CD-HIT (Huang et al. 2010), resulting in a total of 315 sequences used as reference for the primer design (see list in Table S2.1 and taxonomic tree in Figure S2.1). The reference sequences were affiliated to 11 phyla, including *Actinobacteria* (59 sequences), *Bacteroidetes* (22 sequences), *Cyanobacteria* (22 sequences), *Deinococcus-Thermus* (2 sequences), *Ignavibacteriae* (1 sequence), *Firmicutes* (13 sequences), *Gemmatimonadetes* (1 sequence), *Spirochaetes* (16 sequences), *Planctomycetes* (4 sequences), *Proteobacteria* (173 sequences) and *Verrucomicrobia* (2 sequences).

The gene sequences were aligned using MUSCLE (Edgar 2004) and the alignment was manually reviewed by comparison with the aligned translated sequences, using Geneious 6.1.2 (Biomatters, Australia, http://www.geneious.com) and the alignment of the COG3540 group available on NCBI (Conserved Domain Protein Family, http://www.ncbi-nlm.nih.gov/Structure/cdd/) as amino acid reference alignment. The most suitable conserved regions for primer design were identified using PrimerProspector (Walters et al. 2011). Forward and reverse candidate primers were then manually designed to reach the maximum coverage of the reference sequences. Candidate primers were paired to target

an amplicon length of 250 to 500 bp, which represents the best compromise length for next generation sequencing and quantitative PCR studies. They were then tested in silico using De-metaST-BLAST (Gulvik et al. 2012) to identify potential primer pairs with an appropriate product size and coverage of the reference sequences.

2.2.5 Optimization and validation of *phoD*-targeting primers

Candidate primers (21 forward primers and 23 reverse primers) were tested in a gradient PCR using a mixture of soil genomic DNA (Table 2.1, S5-6) as template. PCR reactions were performed in a 25 µl volume containing 1X MyTaq Reaction Buffer (including $MgCl_2$ and dNTPs), 0.5 µM of each primer and 0.6 Units of MyTaq polymerase (Bioline, NSW, Australia) with 1-2 ng DNA as template in an S1000 thermocycler (Bio-Rad Laboratories, CA). The amplification reaction included an initial denaturation step of 5 min at 95°C followed by 35 cycles of a denaturation step of 30 s at 95°C, an annealing step of 30 s at the calculated annealing temperature of each candidate primer pair (gradient of ± 3 degrees) and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. Amplicon size and intensity, and the presence of primer dimers, was assessed visually after electrophoresis on a 1.5% (w/v) agarose gel and staining with ethidium bromide.

Amplicon specificity was evaluated for selected primer pairs by cloning and sequencing. PCR products were ligated at 4°C overnight using pGEM-T® Vector Systems (Promega, Madison, WI) and transformed into chemically competent *E. coli* cells (α-select; F-*deo*R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17 (rk^-, mk^+) *sup*E44 thi-1 *pho*A Δ(*lac*ZYA-*arg*F) U169 Φ80*lac*Z ΔM15 λ^-) following the manufacturer's instructions (Bioline). Restriction fragment length polymorphism (RFLP) profiling of clones with the expected insert size was done using HhaI (Promega; 0.2 U/µl for 3 hours at 37°C), and profiles were visualized by electrophoresis on a 2% (w/v) agarose gel. Representative inserts of unique RFLP profiles were then sequenced (Macrogen Inc., Seoul, Korea). The resulting sequences

were used to evaluate the coverage and specificity of the candidate primer pairs using BLAST (Altschul et al. 1990). Amplicon diversity was examined for three candidate primer pairs by 454 GS-FLX+ sequencing (Roche 454 Life Sciences, Branford, CT) using barcoded primers. Barcoded primer design, sequencing, and initial quality filtering was performed by Research and Testing Laboratory using standard protocols (Lubbock, TX, http://www.researchandtesting.com). Briefly, sequences with a quality score below 25 were trimmed, and chimeras were removed using USEARCH with a clustering at a 4% divergence (Edgar et al. 2011). Denoising was performed using Research and Testing Denoise algorithm, which uses the non-chimeric sequences and the quality scores to create consensus clusters from aligned sequences. Within each cluster, the probability of prevalence of each nucleotide was calculated, and a quality score generated which was then used to remove noise from the dataset.

The primer pair *phoD*-F733 (5'-TGG GAY GAT CAY GAR GT-3') / *phoD*-R1083 (5'-CTG SGC SAK SAC RTT CCA-3') provided the highest *phoD* diversity and coverage (numbers indicate the respective position in the reference *phoD* gene of *Mesorhizobium loti* MAFF303099). *phoD*-F733 anneals to the conserved region that consists of the amino acid residues WDDHE, which contribute to the coordination of two Ca^{2+} cofactors (Rodriguez et al. 2014). In addition, the fragment targeted by *phoD*-F733/R1083 includes two conserved arginine residues. Nevertheless, the variable part of the amplified region also allows a good identification of taxonomy. This primer pair was named PHOD, and used further in this study.

2.2.6 454-sequencing using PHOD- and ALPS-primers

For comparative analysis of PHOD and ALPS primers ALPS-F730/ALPS-R110 (5'-CAG TGG GAC GAC CAC GAG GT-3'/5'-GAG GCC GAT CGG CAT GTC G-3') (Sakurai et al. 2008), *phoD* genes were amplified in pooled duplicate DNA extracts at a concentration of 20 ng μl^{-1} using the PCR conditions described above, with an annealing temperature at

58°C for PHOD and at 57°C for ALPS primers.Samples were then sequenced using 454-GS-FLX+ pyrosequencing (Roche) by Research and Testing Laboratory with a resulting yield between 1642 and 13998 reads per library. The Standard Flowgram Format (SFF) files were submitted to the European Nucleotide Archive (ENA) under the accession number PRJEB7946.

2.2.7 Sequence analysis

Sequencing datasets amplified by PHOD and ALPS primers were analyzed separately using MOTHUR (Schloss et al. 2009). Sequences were analyzed as nucleic acid sequences to keep the maximum information, allow accurate identification, and avoid artifacts due to frameshifts and errors during back-translation (Philippe et al. 2011). After demultiplexing, reads containing ambiguities and mismatches with either the specific primers or the barcode were removed. Reads with an average quality score below 20 were then filtered out. The remaining reads were trimmed at 150 bp and 450 bp as minimum and maximum length, respectively. Across all samples, 92% of the sequences had a length between 320 to 380 bp.

The resulting PHOD- and ALPS-amplified datasets were merged and aligned using the Needleman-Wunsch global alignment algorithm as implemented in MOTHUR, using 6-mers searching and the aligned reference sequences as template. The pairwise distance matrix was calculated from the alignment and sequences were clustered using the k-furthest method as implemented in MOTHUR, with a similarity cutoff at 75% to define the operational taxonomic units (OTUs), as calculated by Tan et al. (2013). OTU matrices were normalized to the smallest library size using the normalized.shared command in MOTHUR to allow comparison between samples. The relative abundance of each OTU was normalized by the total number of reads per sample. The normalized values were then rounded to the nearest integer. Taxonomy assignment was performed using blastn in BLAST+ (Camacho et al. 2009) with a minimum e-value of 1e-8 to retrieve NCBI

sequence identifiers (GI accession number). Subsequently, in-house Perl scripts were used to populate and query a MYSQL database containing the NCBI GI number and taxonomic lineage information (scripts written by Stefan Zoller, Genetic Diversity Centre, ETH Zürich, available on request).

2.2.8 Data analysis

Rarefaction curves were calculated and extrapolated to 5,000 reads to standardize the samples using EstimateS (Version 9, http://purl.oclc.org/estimates). The unconditional variance was used to construct 95% confidence intervals for both interpolated and extrapolated values, which assumes that the reference sample represents a fraction of a larger but unmeasured community. Observed species richness (S_{obs}) based on normalized library size, estimated species richness based on a library size of 5,000 reads (S_{est}) and Chao1 species richness index Chao and Shen (2003) were calculated using EstimateS. Additionally, the Good's coverage (Good 1953) and the alpha diversity estimated by the Shannon-Wiener (H') (Gotelli and Colwell 2011) index were calculated. Student's paired *t*-tests were used to compare S_{obs}, S_{est}, Good's coverage and H' indices between samples.

Similarity between *phoD* bacterial community structures was tested using pairwise *Libshuff* analysis as implemented in MOTHUR with 1,000 iterations (Schloss et al. 2004). Correlation between community composition and environmental variables was tested by redundancy analysis (RDA) followed by an anova on the RDA fit, and a variance partitioning analysis using the vegan package (vegan: Community Ecology Package, **R** package version 2.2-0; http://CRAN.R-project.org/package=vegan) in **R** v.2.15.0 (**R** Core Team, http://www.R-project.org, 2014). Prior to analysis, measured environmental variables (clay and silt content, total C and P, soil pH) were standardized using the z-score method, and nominal variables (vegetation, climate and soil type) were also included.

2.3 Results and Discussion

2.3.1 Taxonomic distribution of *phoD* alkaline phosphatase gene

Our current knowledge of the taxonomic distribution of *phoD* was described based on the IMG/M database. A total of 63 archaeal, 6,469 bacterial and 73 eukaryotic draft or complete genomes containing at least one copy of the *phoD* gene were found. In bacteria, the *phoD* gene was spread across 20 phyla (Figure 2.1a). More than half of the genomes of *Actinobacteria*, *Gemmatimonadetes*, *Spirochaetes* and *Verrucomicrobia* contained at least one copy of the *phoD* gene. Among the *Proteobacteria*, the *phoD* gene occurred in 52, 30 and 34% of the *Alpha-*, *Beta-* and *Gamma-Proteobacteria*. The number of *phoD* copies per genome varied between 1 and 9, but the majority of sequenced genomes (71%) carried only a single copy.

Although *phoD* is widespread across bacterial phyla, it is important to note that the microbial genome sequence database contains almost exclusively the genomes of cultured strains, which creates a general bias in databases (Sait et al. 2002). *Proteobacteria* was the most recurrent phylum in the database, as the *Gamma-Proteobacteria* and more particularly the *Pseudomonas* genus are amongst the most intensively studied taxa (Zappa et al. 2001), and thus the genomes found most frequently in databases. Given the presence of the *phoD* gene in the less represented phyla such as *Chloroflexi*, *Deinococcus-Thermus* and *Planctomycetes*, *phoD*-targeting primers represent an important tool to study these less easily culturable phyla.

Additionally, *phoD* genes were found in archaea, affiliated almost entirely to *Euryarchaeota* (*Halobacteriaceae*), and in eukaryotes, mainly in *Ascomycetes*. Alkaline phosphatase activity in archaea has only rarely been reported, e.g. from extreme environments (Fernandez and Kidney 2007; Wende et al. 2010), while in eukaryotes it has been reported in Basidiomycetes (Šnajdr et al. 2008) and in eukaryotic phytoplanktonic cells (Dyhrman and Ruttenberg 2006), and in mammals it is widely used as an indicator for

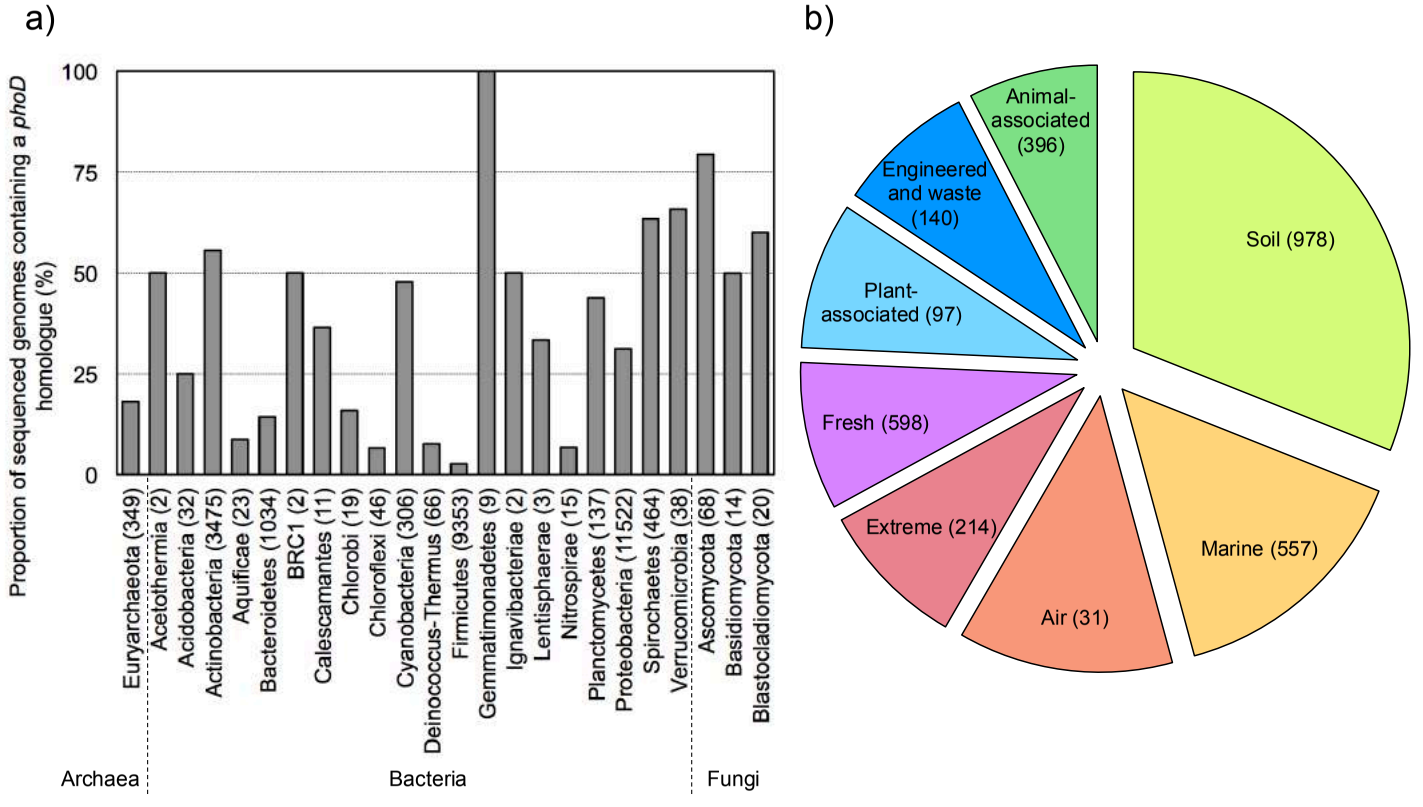


Figure 2.1. Current knowledge of the *phoD* gene in the IMG/M database. (a) Proportion of sequenced genomes containing a *phoD* homologue. Numbers in brackets indicate the total number of sequenced genomes in each phylum. (b) Relative abundance of *phoD* genes in different types of environments (normalized as number of *phoD* counts per number of bases sequenced per metagenome dataset). Numbers in brackets indicate the number of metagenome datasets per environment type.

liver disease (Fernandez and Kidney 2007). However, alkaline phosphatase activity has not previously been associated with the *phoD* gene in these taxa.

2.3.2 Environmental distribution of *phoD* alkaline phosphatase – a meta-analysis

The prevalence of *phoD* in the environment was investigated by analysis of 3,011 available metagenome datasets in the IMG/M database. The *phoD* gene was found in a range of environments (Figure 2.1b), with greatest abundance in soil, followed by marine and air environments.

Metagenomic studies focusing on phosphatases in marine environments have shown that *phoD* and *phoX* are more common than *phoA* in these samples (Sebastián and Ammerman 2009; Luo et al. 2010). The high diversity and relative abundance of the *phoD* gene found in soil metagenomes (Figure 2.1b) suggests that *phoD* may also be particularly relevant in terrestrial ecosystems, though the relative abundance of the three alkaline phosphatase families in soil has not yet been studied on the metagenome level. The fact that organic P represents between 30% and 80% of the total P in grassland and agricultural soils, mainly in the form of diverse phosphomonoesters and diesters (Condron et al. 2005), may promote the diversity of *phoD* in terrestrial ecosystems.

2.3.3 Performance of PHOD and ALPS primers

A key aim of this work was to design a new set of PHOD primers targeting the bacterial *phoD* alkaline phosphatase, for studying the *phoD* bacterial community diversity and composition in soil. The PHOD primers were tested on six soils collected in Australia and Switzerland that represent a range of contrasting soil properties, and the results were compared with those obtained with the same samples using the ALPS primers.

Generally, amplification using PHOD primers resulted in fewer filtered reads than ALPS primers, with 2,309 $\pm$ 1,148 (mean $\pm$ standard deviation) and 7,778 $\pm$ 3,107

Table 2.2. Data obtained with PHOD and ALPS primers based on normalized data.

Primers	Sample	Nb. of filtered reads	Nb. of unique reads	Nb. of reads after normalization	Richness			Good's coverage	H'	Taxonomy				
					S_{obs}	S_{est}	Chao1			Phylum	Class	Order	Family	Genus
PHOD	S1	1915	1763	1088	290	685	684	0.83	4.6	10	15	20	30	37
	S2	2170	1820	963	201	293	303	0.91	3.9	10	14	18	29	39
	S3	3090	2709	1001	227	458	452	0.87	4.2	9	14	18	32	43
	S4	1042	829	1037	148	214	210	0.93	3.8	8	12	13	20	23
	S5	4399	3296	977	191	359	352	0.9	4.2	11	16	21	37	46
	S6	1240	937	1039	199	318	313	0.89	4.2	9	12	14	26	31
ALPS	S1	5958	2097	1017	78	100	97	0.99	3.2	5	8	9	15	18
	S2	12619	3168	998	168	209	290	0.93	3.8	6	10	14	24	32
	S3	3730	1276	1027	139	217	212	0.95	3.8	4	6	7	18	21
	S4	5025	2097	995	123	181	177	0.98	3.1	5	8	12	22	27
	S5	9482	3110	1012	107	143	140	0.97	3.4	4	6	9	14	18
	S6	9854	2038	999	195	238	237	0.98	4	5	7	12	23	29
p-value (Student *t*-test)[1]		**	**	n.s.	*	**	**	***	**	***	***	***	***	***

[1]Student t test: n.s. refers to non-significant, *, ** and *** means statistically significant at *p*-value <0.1, <0.05 and <0.01, respectively.

reads and an average read length of 380 ± 33 bp and 364 ± 35 bp for PHOD- and ALPS-amplified samples, respectively (Table 2.2). The difference in the number of filtered reads per library was directly linked to primer design, more particularly to the degree of degeneracy of the PHOD primers. Increasing degeneracy in primers generally reduces PCR efficiency due to the dilution of each unique primer sequence (Acinas et al. 2005). Degenerate primers increase the risk of unspecific annealing during the PCR reaction, but increase the probability of amplifying yet unknown *phoD* gene sequences by allowing all coding possibilities for an amino acid residue in the nucleic acid sequences (Limansky and Viale 2002). When used appropriately, degenerate primers such as the PHOD primers represent a great advantage in studies on genetic diversity by targeting known and unknown sequences in environmental samples (Menzel et al. 2011).

By filtering out redundant sequences, the number of reads decreased remarkably in the ALPS-amplified samples, leading to more similar numbers of unique reads for the two sets of primers which averaged 1,893 ± 885 bp (mean ± standard deviation) and 2,297 ± 659 bp for PHOD- and ALPS-amplified samples, respectively. This showed that ALPS-amplified samples consisted of a greater number of redundant reads than PHOD-amplified samples. Finally, normalization of the library size in order to compare the two primer sets resulted in an average library size of 1,013 ± 31 bp. Our results suggest that the ALPS primers target a narrow spectrum of sequences which represent a large fraction of the reads after amplification.

2.3.4 Species richness and alpha diversity of the *phoD* gene in six soils

Amplification with PHOD primers revealed two-fold variation in species richness among the six samples (Table 2.2). S_{obs} was lowest in S4 and highest in S1, with 148 and 290 OTUs, respectively. Chao1 and S_{est} indices, derived from the rarefaction curves, showed a similar trend. The difference in species richness between samples is well illustrated by the rarefaction curves (Figure 2.2a). The rarefaction curve of S1 had the steepest slope,

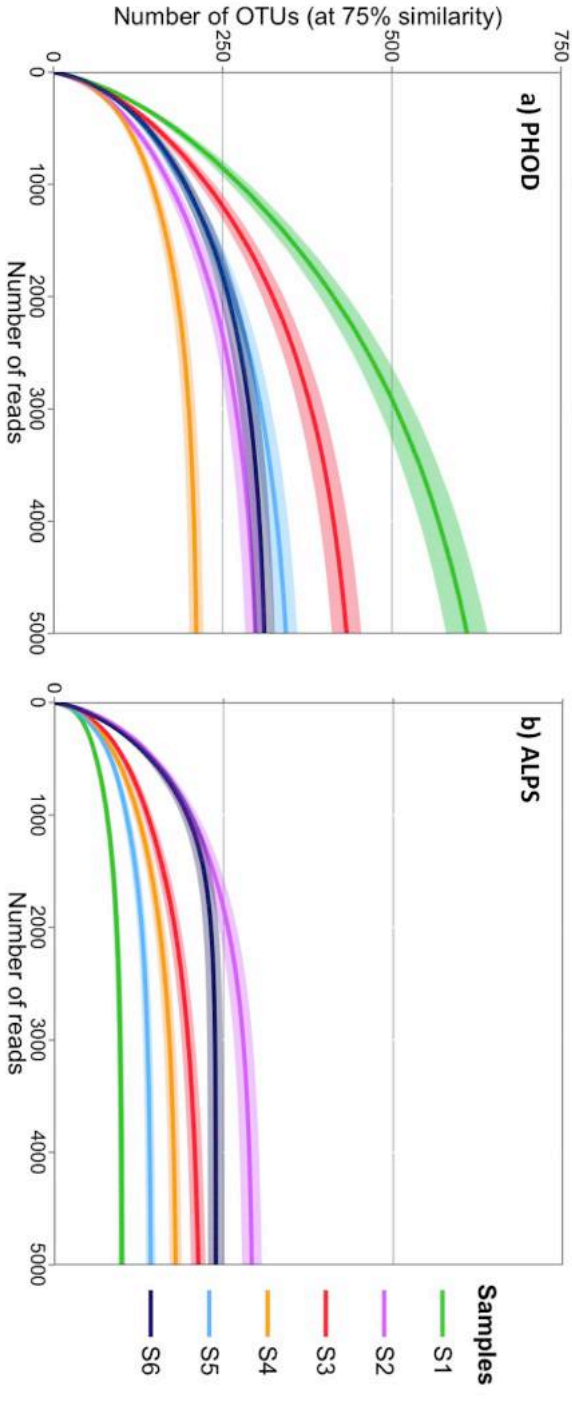


Figure 2.2. Rarefaction curves of the samples S1-6 amplified by PHOD (a) and ALPS (b) primers extrapolated to 5000 reads with 95% confidence intervals.

showing the greatest increase of OTUs with the number of reads, while that of S4 reached the asymptote with the fewest reads (ca. 3,000 reads).

Compared to amplification with PHOD primers, amplification with ALPS primers resulted in significantly lower species richness and alpha diversity (Table 2.2). In ALPS-amplified samples, the rarefaction curves always reached the asymptote with fewer reads than in the corresponding PHOD-amplified samples (Figure 2.2a and b). The rarefaction curves of S1 when amplified using PHOD and ALPS primers contrasted particularly strongly, leading to a seven-fold difference in Chao1 and S_{est}. Likewise, H' was always greater in PHOD- than in ALPS-amplified samples. This suggests that PHOD primers target a broader diversity of *phoD*-bearing bacteria than ALPS primers.

Using ALPS primers, Tan et al. (2013) found between 450 and 548 OTUs in soils fertilized with zero, medium or high P input, with a sequencing depth between 14,279 and 16,140 reads. In contrast, Fraser et al. (2015a) reported lower numbers which are in the same range as in the six soils analysed in this study. They found between 137 and 163 OTUs in soils from organic and conventional cropping systems and prairie, with a sequencing depth of 11,537 to 54,468 reads. Thus, the number of OTUs seems to be quite variable between studies and/or soils. By applying both primers on the same soils, we found that PHOD primers targeted a larger species spectrum than ALPS primers.

2.3.5 Dominant phyla harboring *phoD* in six soils

Taxonomy was assigned to most sequences using BLAST+ (Camacho et al. 2009) (Figure 2.3 and Table S2.2). A remainder of 5,052 reads could not be assigned taxonomic identity, representing between 0.1 and 22% of the total filtered read number. In theory, the primers could amplify *phoD* also in archaea and eukaryotes, as *phoD* has been found in several archaeal and eukaryotic species in the IMG/M database. In the six soils studied here, both ALPS and PHOD primers amplified phoD from bacteria only, based on identification using BLAST+.

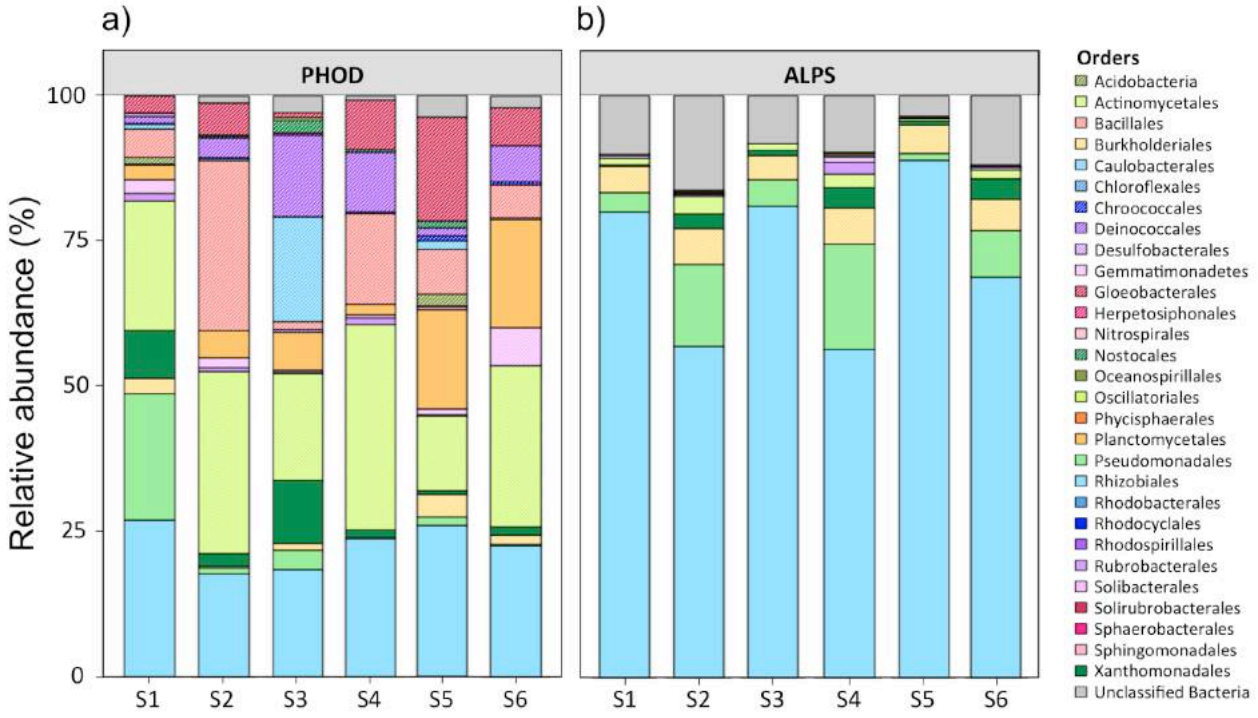


Figure 2.3. Relative abundances in percent of the total community at the order level in the samples S1-6 amplified by PHOD (a) and ALPS primers (b).

PHOD primers targeted *phoD* genes from 13 phyla (*Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Nitrospirae*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes* and *Verrucomicrobia*). They covered 22 classes, 38 orders, 71 families and 113 genera. The dominant orders were *Actinomycetales* (13-35%), *Bacillales* (1-29%), *Gloeobacteriales* (1-18%), *Rhizobiales* (18-27%) and *Pseudomonadales* (0-22%). *Libshuff* analysis showed that the *phoD* bacterial communities in the different soils were significantly different from each other (p-value <0.001). S1 was characterized by 25% of *Pseudomonadales* and 10% of *Xanthomonadales*. The highest relative abundance of *Bacillales* (29%) was found in S2. S3 was composed by a particularly high abundance of *Caulobacteriales* (19%), *Deinococcales* (14%) and *Xanthomonadales* (11%). *Planctomycetes* were especially abundant in S4 and S6 with 18 and 19%, respectively, while S5 showed a high abundance of *Gloeobacteriales* (18%). ALPS primers amplified *phoD* genes from 6 phyla (*Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Gemmatimonadetes* and *Proteobacteria*). In more detail, ALPS primers covered 13 classes, 22 orders, 42 families and 64 genera. The most prevalent class was *Alpha-Proteobacteria* (55-92%). *Rhizobiales* was the dominant taxon in this class, with an overrepresentation of *Methylobacterium* sp., which represented 60-95% of the abundance of *Rhizobiales*. *Libshuff* analysis showed that the structures of the *phoD* bacterial communities in the different samples were also significantly different from each other (p-value <0.001).

This taxonomy analysis highlights the fact that the *phoD* gene is widespread across phyla and that the PHOD primers covered the *phoD* diversity well. PHOD primers targeted *phoD* genes in 13 out of the 20 phyla known to carry the *phoD* gene, based on the IMG/M database. PHOD primers captured a particularly large diversity of *Actinobacteria*, including the common soil genera *Actinomyces, Arthrobacter, Keinococcus, Kitasatospora, Micrococcus* and *Streptosporangium* (Ward and Bora 2008), and of *Proteobacteria* including *Azorhizobium, Rhodospirillum, Caulobacter, Geobacter* and *Variovorax* (Nacke et al. 2011). Both *Actinobacteria* and *Proteobacteria* are known to

be important for mineralization of soil organic matter and in composting processes (Yu et al. 2007; Danon et al. 2008). Our sequencing results for soils, in accordance with the IMG/M analysis, show that a greater diversity of microorganisms than previously thought contributes to organic P mineralization by secreting PhoD.

PHOD primers amplified many sequences belonging to phyla with low abundance in the IMG/M database. These sequences were affiliated to the phyla (e.g. *Deinobacter* sp.), *Nitrospirae* (e.g. *Nitrospira* sp.), *Spirochaetes* (e.g. *Spirochaeta* sp.), *Planctomycetes* (e.g. *Isophaera* sp. and *Planctomyes* sp.) and *Verrucomicrobia* (e.g. *Opitutus* sp.). The ALPS primers did not amplify *phoD* genes from these phyla. Moreover, compared to the PHOD primers, the ALPS primers failed to detect *phoD* genes from many genera, including e.g. *Anabaena*, *Chroococcidiopsis*, and *Chroococcus* belonging to the *Cyanobacteria*. Our results support the conclusion of Tan et al. (2013) that the ALPS primers have an amplification bias, restraining the amplification to a limited number of microbial taxa and over-representing *Alpha-Proteobacteria*, probably because of the few sequences used to design the primers (7 sequences from 4 phyla used, compared with 315 sequences from 11 phyla used here for PHOD primer design).

2.3.6 Soil pH is the main driver of the *phoD* bacterial community

Redundancy analysis (RDA) of PHOD-amplified samples indicated that 49.1% of the variation was explained by the two main RDA components (Figure 2.4). Variance partitioning analysis showed that soil pH explained 23.7% and total P 18.3% of the variance among the communities. However, soil pH was the only environmental variable that was significantly correlated with the distribution of the samples (p-value = 0.03). The most divergent samples along the first RDA component axis were S1 and S4. The observed differences between these samples are likely due to the very contrasting soil and environmental properties between the sampling sites. S1 was sampled in an oceanic and temperate climatic region with dense vegetation, while S4 was collected in a hot

semi-arid climatic region with only scattered vegetation, where lower soil microbial biomass and diversity is expected (Bachar et al. 2010). S1 and S4 also exhibited the biggest difference in soil pH, which is regarded as the main environmental driving force that affects total microbial communities and activities (Fierer and Jackson 2006; Lauber et al. 2009). Soil pH has previously been observed to be an important driver of *phoD* bacterial community in the rhizosphere of wheat grown in different soils (Wang et al. 2012b). Phosphatase activity can respond to changes in soil pH within days, e.g. after a lime treatment in agricultural soils (Dick et al. 2000). The second RDA component was mainly linked to total P. The *phoD* communities of S5 and S6 clustered together along the second component axis, probably because these two samples were both collected in Switzerland and had high total carbon and other similar soil properties. In contrast, S1, S3 and S4 had low total C and P values.

Previous studies using the ALPS primers have reported an effect of the application of organic and conventional fertilizers, crop management, vegetation and pH (Wang et al. 2012a; Wang et al. 2012b; Chhabra et al. 2013; Tan et al. 2013; Jorquera et al. 2014; Fraser et al. 2015a; Fraser et al. 2015b). Plant community has been reported to have an impact on *phoD* diversity and community structure in monocultures (Wang et al. 2012a; Wang et al. 2012b). P fertilization has been reported to either increase (Tan et al. 2013) or to reduce (Chhabra et al. 2013) the diversity of the *phoD* gene. Jorquera et al. (2014) observed that P fertilization alone did not affect *phoD* bacterial community structure in a Chilean Andisol pasture, while combined N and P fertilization did change the *phoD* bacterial community structure. While all these studies have provided some insights into the environmental drivers affecting *phoD* bacterial communities, they need to be interpreted with caution due to the amplification bias of the ALPS primers towards *Alpha-Proteobacteria* described above. PHOD primers should now be applied to a wider range of soils to verify whether pH is the main driver of the *phoD* bacterial community.

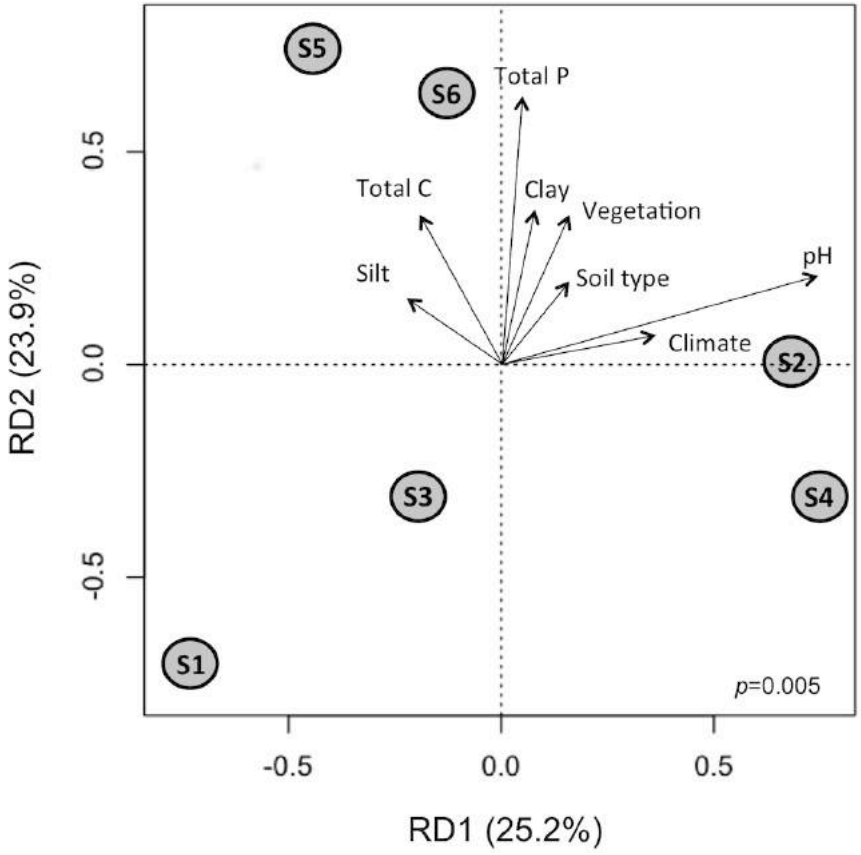


Figure 2.4. Redundancy analysis of *phoD* bacterial community of the samples S1-6 amplified by PHOD primers with the environmental variables clay and silt content, total C and P, soil type, climate, vegetation and soil pH. The significance of the model is indicated in the bottom right corner. Note that soil pH was the unique environmental variable that was significantly correlated with the *phoD* bacterial community (p-value = 0.03.).

2.4 Conclusion

Evaluation of metagenomic datasets revealed that the *phoD* gene is primarily found in bacteria and is spread across 20 bacterial phyla. *phoD* has been found to be ubiquitous in the environment with terrestrial ecosystem metagenomes containing the highest relative

abundance of *phoD*. The newly designed PHOD primers reported here covered the large diversity of the *phoD* gene better than previously published primers and amplified sequences affiliated to 13 bacterial phyla. The most prevalent *phoD* genes identified in six diverse soils from Europe and Australia were affiliated with the orders *Actinomycetales, Bacillales, Gloeobacteriales, Rhizobiales* and *Pseudomonadales*. Soil pH was found to be the main environmental driver affecting the *phoD* bacterial community. PHOD primers can be used as a tool to study *phoD* bacterial community diversity and composition, and to identify and quantify microorganisms that carry and express *phoD* in the environment.

2.5 Acknowledgments

The authors would like to thank Dr. Stefan Zoller for the Perl scripts for taxonomic analysis and the Genetic Diversity Center (Zürich, Switzerland) for technical assistance. We also acknowledge Agroscope (Switzerland) and the New South Wales Department of Primary Industry (NSW, Australia) for access to the sampling sites. This work was supported by the Swiss National Science Foundation (SNF) and by a research grant from the University of Sydney.

2.6 Supplementary material

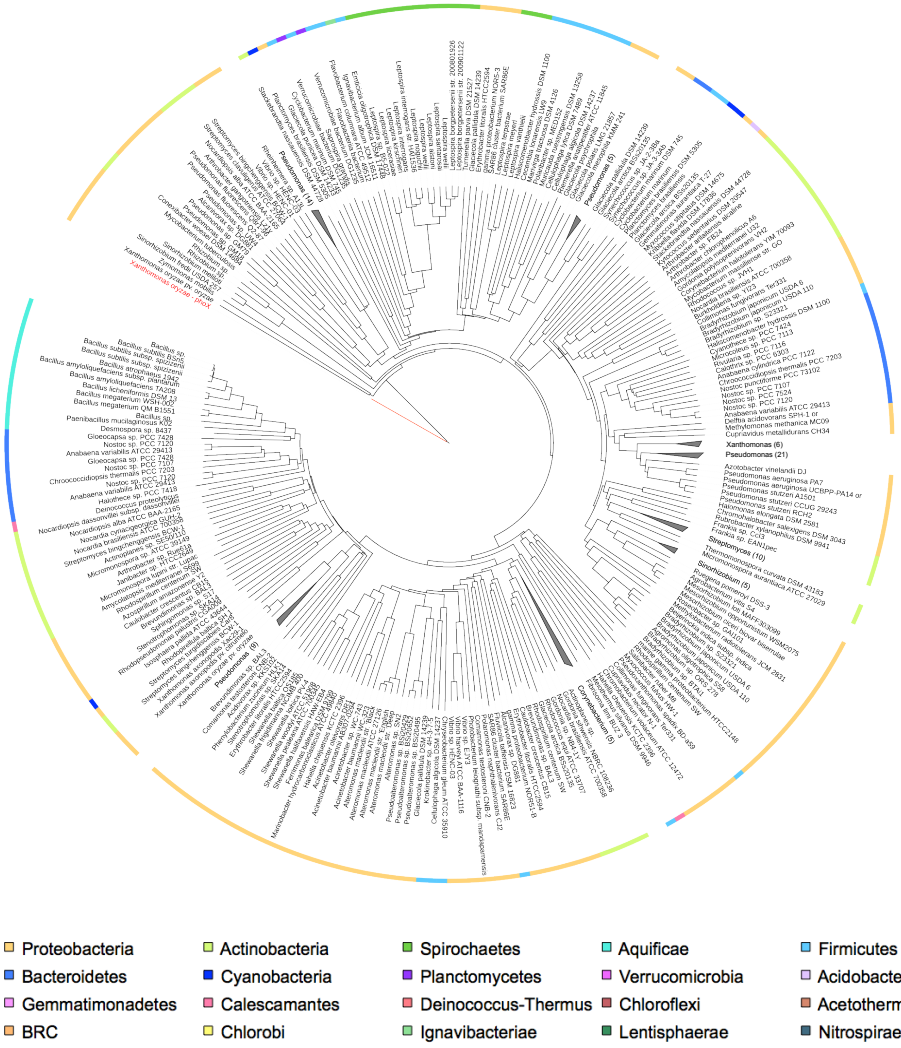


Figure S2.1. Neighbor-joining tree based on the amino acid sequences of the full length of the *phoD* genes. The tree was rooted using the sequence of the *phoX* alkaline phosphatase gene of *Xanthomonas oryzae* (in red).

Table S2.1. Taxonomic summary showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples S1-6 amplified by PHOD and ALPS primers.

	Number of reads											
	PHOD						ALPS					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
1. Acidobacteria	38	0	0	0	70	0	0	0	0	0	0	0
1.1. Acidobacteria	16	0	0	0	70	0	0	0	0	0	0	0
1.1.1. Acidobacteria	16	0	0	0	70	0	0	0	0	0	0	0
1.1.1.1. Acidobacteriaceae	16	0	0	0	70	0	0	0	0	0	0	0
1.1.1.1.1. Acidobacterium	15	0	0	0	66	0	0	0	0	0	0	0
1.1.1.1.2. Granulicella	1	0	0	0	4	0	0	0	0	0	0	0
1.2. Solibacteres	7	0	0	0	0	0	1	0	1	0	0	0
1.2.1. Solibacterales	7	0	0	0	0	0	1	0	1	0	0	0
1.2.1.1. Solibacteraceae	7	0	0	0	0	0	1	0	1	0	0	0
1.2.1.1.1. Solibacter	7	0	0	0	0	0	1	0	1	0	0	0
1.3. Unclassified Acidobacteria	15	0	0	0	0	0	0	0	0	0	0	0
2. Actinobacteria	325	611	392	343	443	285	26	101	18	71	21	59
2.1. Actinomycetales	307	598	388	331	440	285	26	93	18	39	15	59
2.1.1. Actinoplanaceae	198	434	209	109	211	133	25	62	10	36	15	27
2.1.1 1. Actinoplanes	106	198	3	0	24	11	0	0	0	0	1	3
2.1.1.2. Kitasatospora	1	0	3	1	0	0	0	0	0	0	0	0
2.1.1.3. Micromonospora	1	22	2	6	36	13	0	0	6	0	0	0
2.1.1.4. Microstreptospora	26	12	61	0	44	16	0	3	0	5	0	9
2.1.1.5. Streptomyces	64	201	127	102	94	93	25	59	4	31	14	15
2.1.1.6. Streptosporangium	0	1	13	0	13	0	0	0	0	0	0	0
2.1.2. Actinosynnemataceae	17	31	29	157	0	28	1	2	1	0	0	28
2.1.2.1. Actinomyces	0	2	0	0	0	22	0	0	0	0	0	0
2.1.2.2. Amycolatopsis	4	19	5	155	0	4	0	0	1	0	0	1
2.1.2.3. Kutzneria	0	0	0	0	0	0	0	0	0	0	0	27
2.1.2.4. Saccharothrix	13	10	20	2	0	2	0	2	0	0	0	0
2.1.2.5. Thermobispora	0	0	4	0	0	0	1	0	0	0	0	0
2.1.3. Beutenbergiaceae	6	5	0	0	28	48	0	0	0	0	0	0
2.1.3.1. Beutenbergia	6	5	0	0	28	48	0	0	0	0	0	0
2.1.4. Cellulomonadaceae	0	0	0	0	0	0	0	0	1	0	0	0
2.1.4.1. Cellulomonas	0	0	0	0	0	0	0	0	1	0	0	0
2.1.5. Corynebacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0
2.1.5.1. Corynebacterium	0	0	0	0	0	0	0	0	0	0	0	0
2.1.6. Frankiaceae	14	0	62	2	34	0	0	0	0	0	0	1
2.1.6.1. Frankia	14	0	62	2	34	0	0	0	0	0	0	1
2.1.7. Geodermatophilaceae	0	0	0	0	0	0	0	0	1	1	0	1
2.1.7.1. Blastococcus	0	0	0	0	0	0	0	0	0	1	0	0
2.1.7.2. Geodermatophilus	0	0	0	0	0	0	0	0	1	0	0	0
2.1.7.3. Modestibacter	0	0	0	0	0	0	0	0	0	0	0	1
2.1.8. Glycomycetaceae	0	0	0	0	0	1	0	8	0	0	0	0
2.1.8.1. Stackebrandtia	0	0	0	0	0	1	0	8	0	0	0	0
2.1.9. Gordoniaceae	0	0	0	0	0	0	0	0	0	0	0	0
2.1.9.1. Gordana	0	0	0	0	0	0	0	0	0	0	0	0
2.1.10. Kineosporiaceae	7	13	5	3	32	11	0	0	0	0	0	0
2.1.10.1. Kineococcus	7	13	5	3	32	11	0	0	0	0	0	0
2.1.11. Micrococcaceae	6	15	40	25	2	2	0	0	0	0	0	0
2.1.11.1. Arthrobacter	1	0	21	0	1	0	0	0	0	0	0	0
2.1.11.2. Micrococcus	5	15	19	25	1	2	0	0	0	0	0	0
2.1.12. Mycobacteriaceae	0	0	0	0	3	1	0	0	0	0	0	2
2.1.12.1. Mycobacterium	0	0	0	0	3	1	0	0	0	0	0	2
2.1.13. Nocardiaceae	0	5	0	0	0	0	0	0	0	0	0	0
2.1.13.1. Micropolyspora	0	0	0	0	0	0	0	0	0	0	0	0
2.1.13.2. Rhodococcus	0	5	0	0	0	0	0	0	0	0	0	0

Table S2.1. Taxonomic summary showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples S1-6 amplified by PHOD and ALPS primers. (Continued)

	Number of reads											
	PHOD						ALPS					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
2.1.14. Nocardioidaceae	59	87	34	35	127	61	0	14	5	0	0	0
2.1.14.1. Kribella	59	87	28	34	126	61	0	14	5	0	0	0
2.1.14.2. Nocardia	0	0	6	1	1	0	0	0	0	0	0	0
2.1.15. Propionibacteriaceae	0	0	0	0	0	0	0	0	0	2	0	0
2.1.15.1. Microlunatus	0	0	0	0	0	0	0	0	0	2	0	0
2.1.16. Thermomonosporaceae	0	8	9	0	2	0	0	7	0	0	0	0
2.1.16.1. Thermomonospora	0	8	9	0	2	0	0	7	0	0	0	0
2.1.17. Tsukamurellaceae	0	0	0	0	1	0	0	0	0	0	0	0
2.1.17.1. Tsukamurella	0	0	0	0	1	0	0	0	0	0	0	0
2.2. Rubrobacterales	18	13	4	12	3	0	0	8	0	32	0	0
2.2.1. Rubrobacteraceae	18	13	4	12	3	0	0	8	0	32	0	0
2.2.1.1. Rubrobacter	18	13	4	12	3	0	0	8	0	32	0	0
2.3. Solirubrobacterales	0	0	0	0	0	0	0	0	0	0	6	0
2.3.1. Conexibacteraceae	0	0	0	0	0	0	0	0	0	0	6	0
2.3.1.1. Conexibacter	0	0	0	0	0	0	0	0	0	0	6	0
3. Bacteroidetes	0	0	0	1	4	0	2	3	0	0	0	0
3.1. Sphingobacteriia	0	0	0	1	4	0	2	3	0	0	0	0
3.1.1. Chitinophagaceae	0	0	0	1	4	0	2	3	0	0	0	0
3.1.1.1. Niastella	0	0	0	1	4	0	2	3	0	0	0	0
4. Chloroflexi	4	0	6	0	0	0	0	2	9	0	0	0
4.1. Chloroflexi	2	0	6	0	0	0	0	1	0	0	0	0
4.1.1. Chloroflexaceae	2	0	0	0	0	0	0	1	0	0	0	0
4.1.1.1. Roseiflexaceae	2	0	0	0	0	0	0	1	0	0	0	0
4.1.1.1.1.Roseiflexus	2	0	0	0	0	0	0	1	0	0	0	0
4.1.2. Herpetosiphonales	0	0	6	0	0	0	0	0	0	0	0	0
4.1.2.1. Herpetosiphonaceae	0	0	6	0	0	0	0	0	0	0	0	0
4.1.2.1.1. Herpetosiphon	0	0	6	0	0	0	0	0	0	0	0	0
4.2. Thermomicrobia	2	0	0	0	0	0	0	1	9	0	0	0
4.2.1. Sphaerobacterales	2	0	0	0	0	0	0	1	0	0	0	0
4.2.1.1. Sphaerobacteraceae	2	0	0	0	0	0	0	1	0	0	0	0
4.2.1.1.1. Sphaerobacter	2	0	0	0	0	0	0	1	0	0	0	0
5. Cyanobacteria	39	116	66	88	673	72	0	0	0	0	0	5
5.1. Cyanophyceae	0	6	47	6	43	0	0	0	0	0	0	0
5.1.1. Chroococcales	0	0	1	0	0	0	0	0	0	0	0	0
5.1.1.1. Xenococcaceae	0	0	1	0	0	0	0	0	0	0	0	0
5.1.1.1.1. Chroococcidiopsis	0	0	0	0	0	0	0	0	0	0	0	0
5.1.1.1.2. Pleurocapsa	0	0	1	0	0	0	0	0	0	0	0	0
5.1.2. Nostocales	0	6	46	6	43	0	0	0	0	0	0	0
5.1.2.1. Nostocaceae	0	6	46	6	43	0	0	0	0	0	0	0
5.1.2.1.1. Amorphonostoc	0	6	46	5	0	0	0	0	0	0	0	0
5.1.2.1.2. Anabaena	0	0	0	1	43	0	0	0	0	0	0	0
5.2. Gloeobacteria	39	107	19	80	605	67	0	0	0	0	0	5
5.2.1. Gloeobacterales	39	107	19	80	605	67	0	0	0	0	0	5
5.2.1.1. Gloeobacter	39	107	19	80	605	67	0	0	0	0	0	5
5.2.1.1.1. Gloeobacter	39	107	19	80	605	67	0	0	0	0	0	5
5.3. Oscillatoriophycideae	0	3	0	2	25	5	0	0	0	0	0	0
5.3.1. Chroococcales	0	3	0	2	25	5	0	0	0	0	0	0
5.3.1.1. Chroococcales	0	3	0	2	25	5	0	0	0	0	0	0
5.3.1.1.1. Chroococcus	0	2	0	2	25	5	0	0	0	0	0	0
5.2.1.1.2. Cyanothece	0	1	0	0	0	0	0	0	0	0	0	0
5.3.2. Oscillatoriales	0	0	11	0	0	0	0	0	0	0	0	0
5.3.2.1. Oscillatoriales	0	0	11	0	0	0	0	0	0	0	0	0
5.3.2.1.1. Microcoleus	0	0	11	0	0	0	0	0	0	0	0	0

Table S2.1. Taxonomic summary showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples S1-6 amplified by PHOD and ALPS primers. (Continued)

	Number of reads											
	PHOD						ALPS					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
6. Deinococcus-Thermus	16	63	296	95	47	64	0	0	0	0	0	0
6.1. Hadobacteria	16	63	296	95	47	64	0	0	0	0	0	0
6.1.1. Deinococcales	16	63	296	95	47	64	0	0	0	0	0	0
6.1.1.1. Deinococcaceae	16	63	296	95	47	64	0	0	0	0	0	0
6.1.1.1.1. Deinobacter	16	63	296	95	47	64	0	0	0	0	0	0
6.1.1.2. Trueperaceae	0	0	0	0	0	0	0	0	0	0	0	0
6.1.1.2.1. Truepera	0	0	0	0	0	0	0	0	0	0	0	0
7. Firmicutes	67	561	30	145	263	58	0	0	0	0	0	0
7.1. Bacilli	67	561	30	145	263	58	0	0	17	0	0	0
7.1.1. Bacillales	67	561	30	145	263	58	0	0	0	0	0	0
7.1.1.1. Bacillaceae	67	561	30	145	263	58	0	0	0	0	0	0
7.1.1.1.1. Bacillus	67	549	25	145	263	58	0	0	0	0	0	0
7.1.1.1.2. Geobacillus	0	12	5	0	0	0	0	0	0	0	0	0
7.2. Clostridia	0	0	0	0	0	0	0	0	0	0	0	0
7.2.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	0
7.2.1.1. Peptococcaceae	0	0	0	0	0	0	0	0	0	0	0	0
7.2.1.1.1. Desulfitobacterium	0	0	0	0	0	0	0	0	0	0	0	0
8. Gemmatimonadetes	33	33	6	4	37	67	12	0	0	11	5	21
8.1. Gemmatimonadetes	29	33	6	4	34	67	0	0	0	0	0	3
8.1.1. Gemmantimonadales	29	33	6	4	34	67	0	0	0	0	0	3
8.1.1.1. Gemmantimonadaceae	29	33	6	4	34	67	0	0	0	0	0	3
8.1.1.1.1. Gemmatimonas	29	33	6	4	34	67	0	0	0	0	0	3
8.2. Unclassified Gemmatimonadetes	4	0	0	0	3	0	12	0	0	11	5	18
9. Nitrospirae	0	2	0	0	0	0	0	0	0	0	0	0
9.1. Nitrospira	0	2	0	0	0	0	0	0	0	0	0	0
9.1.1. Nitrospirales	0	2	0	0	0	0	0	0	0	0	0	0
9.1.1.1. Nitrospiraceae	0	2	0	0	0	0	0	0	0	0	0	0
9.1.1.1.1. Nitrospira	0	2	0	0	0	0	0	0	0	0	0	0
10. Planctomycetes	34	87	139	18	582	190	0	2	139	5	0	0
10.1. Planctomycetacia	34	87	139	18	582	190	0	0	0	0	0	0
10.1.1. Planctomycetales	34	87	139	18	582	190	0	0	0	0	0	0
10.1.1.1. Planctomycetaceae	34	87	139	18	582	190	0	0	0	0	0	0
10.1.1.1.1. Isosphaera	6	1	3	0	104	9	0	0	0	0	0	0
10.1.1.1.2. Pirella	22	41	19	0	25	13	0	0	0	0	0	0
10.1.1.1.3. Planctomyces	0	6	0	0	0	0	0	0	0	0	0	0
10.1.1.1.4. Rhodopirellula	0	0	0	0	0	0	0	0	0	0	0	0
10.1.1.1.5. Singulisphaera	6	39	117	18	453	168	0	0	0	5	0	0
11. Proteobacteria	642	412	1097	239	1164	269	1878	2477	1592	1537	3405	3350
11.1. Alphaproteobacteria	189	343	772	226	960	236	1707	1760	1423	900	3164	2690
11.1.1. Caulobacter	11	5	381	2	52	0	0	0	0	0	0	0
11.1.1.1. Caulobacteraceae	11	5	381	2	52	0	0	0	0	0	0	0
11.1.1.1.1. Asticcacaulis	0	0	0	0	0	0	0	0	0	0	0	0
11.1.1.1.2. Brevundimonas	0	0	0	0	0	0	0	0	0	0	0	0
11.1.1.1.3. Caulobacter	0	0	0	0	0	0	0	0	0	0	0	0
11.1.1.1.4. Phenylobacterium	11	5	381	2	52	0	0	0	0	0	0	0
11.1.2. Rhizobiales	178	338	391	224	891	232	1707	1760	1423	892	3162	2685
11.1.2.1. Beijerinckiaceae	0	1	0	0	0	5	0	0	0	2	0	1
11.1.2.1.1. Beijerinckia	0	1	0	0	0	5	0	0	0	2	0	1
11.1.2.2. Bradyrhizobiaceae	28	57	220	44	105	42	666	594	219	115	118	247
11.1.2.2.1. Bradyrhizobium	28	31	203	44	105	27	666	594	219	115	118	247
11.1.2.2.2. Oligotropha	0	0	15	0	0	0	0	0	0	0	0	0

Table S2.1. Taxonomic summary showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples S1-6 amplified by PHOD and ALPS primers. (Continued)

	Number of reads											
	PHOD						ALPS					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
11.1.2.2.3. Rhodopseudomonas	0	26	2	0	0	3	0	0	0	0	0	0
11.1.2.3. Methylobacteriaceae	39	11	52	180	466	59	1020	1056	1137	670	3000	2398
11.1.2.3.1. Methylobacterium	39	11	52	180	466	59	1020	1056	1137	670	3000	2398
11.1.2.4. Methylocystaceae	0	4	1	0	0	23	0	0	0	0	0	0
11.1.2.4.1. Methylocystis	0	4	1	0	0	23	0	0	0	0	0	0
11.1.2.5. Phyllobacteriaceae	0	25	15	0	209	55	0	42	10	41	41	33
11.1.2.5.1. Mesorhizobium	0	25	15	0	209	55	0	42	10	41	41	33
11.1.2.6. Rhizobiaceae	5	42	95	0	87	34	21	68	57	64	2	3
11.1.2.6.1. Agrobacterium	0	19	0	0	23	14	0	13	0	0	0	0
11.1.2.6.2. Pelagibacterium	0	0	36	0	3	0	0	0	0	0	0	0
11.1.2.6.3. Rhizobium	5	16	37	0	55	13	20	40	57	0	2	0
11.1.2.6.4. Sinorhizobium	0	7	22	0	6	7	1	15	0	64	0	3
11.1.2.7. Xanthobacteraceae	0	0	5	0	0	3	0	0	0	0	0	0
11.1.2.7.1. Azorhizobium	0	0	1	0	0	2	0	0	0	0	0	0
11.1.2.7.2. Starkeya	0	0	4	0	0	1	0	0	0	0	0	0
11.1.3. Rhodobacterales	0	0	0	0	0	0	0	0	0	6	2	1
11.1.3.1. Rhodobacteraceae	0	0	0	0	0	0	0	0	0	6	2	1
11.1.3.1.1. Ketogulonicigenium	0	0	0	0	0	0	0	0	0	6	2	1
11.1.4. Rhodospirillales	0	0	0	0	0	0	0	0	0	0	0	3
11.1.4.1. Gluconobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0
11.1.4.1.1. Acidiphilium	0	0	0	0	0	0	0	0	0	0	0	0
11.1.4.1.2. Gluconobacter	0	0	0	0	0	0	0	0	0	0	0	0
11.1.4.2. Rhodospirillaceae	0	0	0	0	0	0	0	0	0	0	0	3
11.1.4.2.1. Azospirillum	0	0	0	0	0	0	0	0	0	0	0	0
11.1.4.2.2. Rhodospirillum	2	2	2	2	2	2	2	2	2	2	2	2
11.1.5. Sphingomonadales	0	0	0	0	17	4	0	0	0	2	0	1
11.1.5.1. Sphingomonadaceae	0	0	0	0	17	4	0	0	0	2	0	1
11.1.5.1.1. Sphingobium	0	0	0	0	17	0	0	0	0	0	0	0
11.1.5.1.2. Sphingomonas	0	0	0	0	0	4	0	0	0	1	0	1
11.1.5.1.3. Sphingopyxis	0	0	0	0	0	0	0	0	0	1	0	0
11.2. Betaproteobacteria	36	6	26	1	132	18	97	192	72	98	173	206
11.2.1. Burkholderiales	36	6	26	1	132	18	97	192	72	98	173	206
11.2.1.1. Alcaligenaceae	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.1.1. Achromobacter	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.1.2. Bordetella	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.2. Burkholderiaceae	0	0	7	0	24	1	19	183	68	71	141	177
11.2.1.2.1. Burkholderia	0	0	0	0	2	0	2	18	1	0	8	2
11.2.1.2.2. Cupravidus	0	0	2	0	0	1	17	99	41	34	102	117
11.2.1.2.3. Ralstonia	0	0	5	0	22	0	0	66	26	37	31	58
11.2.1.3. Comamonadaceae	36	0	19	1	106	1	1	9	3	18	32	28
11.2.1.3.1. Acidivorax	1	0	8	1	0	0	0	0	0	0	0	0
11.2.1.3.2. Albidoferax	0	0	0	0	10	1	0	0	0	0	0	0
11.2.1.3.3. Alicycliphilus	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.3.4. Curvibacter	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.3.5. Delftia	0	0	0	0	0	0	0	6	3	18	32	28
11.2.1.3.6. Ramlibacter	0	0	0	0	62	0	1	0	0	0	0	0
11.2.1.3.7. Variovorax	35	0	11	0	34	0	0	0	0	0	0	0
11.2.1.3.8. Verminephrobacter	0	0	0	0	0	0	0	3	0	0	0	0
11.2.1.4. Oxalobacteraceae	0	5	0	0	1	0	76	0	1	9	0	1
11.2.1.4.1. Collimonas	0	0	0	0	1	0	75	0	0	8	0	1
11.2.1.4.2. Herbaspirillum	0	5	0	0	0	0	1	0	1	1	0	0
11.2.1.5. Unclassified Burkholderiales	0	1	0	0	1	16	1	0	0	0	0	0
11.2.1.5.1. Leptothrix	0	0	0	0	0	0	0	0	0	0	0	0
11.2.1.5.2. Methylibium	0	1	0	0	1	16	1	0	0	0	0	0

Table S2.1. Taxonomic summary showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples S1-6 amplified by PHOD and ALPS primers. (Continued)

	Number of reads											
	PHOD						ALPS					
	S1	S2	S3	S4	S5	S6	S1	S2	S3	S4	S5	S6
11.2.2. Neisseriales	0	0	0	0	0	0	0	0	0	0	0	0
11.2.2.1. Chromobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0
11.2.2.1.1. Pseudogulbenkiania	0	0	0	0	0	0	0	0	0	0	0	0
11.2.3. Rhodocyclales	0	0	0	0	0	0	0	5	0	0	0	0
11.2.3.1. Rhodocyclaceae	0	0	0	0	0	0	0	5	0	0	0	0
11.2.3.1.1. Azoarcus	0	0	0	0	0	0	0	5	0	0	0	0
11.3. Gammaproteobacteria	413	63	299	12	71	15	73	523	97	538	68	454
11.3.1.Chromatiaceae	0	0	0	0	0	0	0	0	0	0	0	0
11.3.1.1. Chromatiaceae	0	0	0	0	0	0	0	0	0	0	0	0
11.3.1.1.1. Marichromatium	0	0	0	0	0	0	0	0	0	0	0	0
11.3.2. Oceanospirillales	0	2	0	0	0	0	0	8	0	0	0	0
11.3.2.1. Halomonadaceae	0	2	0	0	0	0	0	8	0	0	0	0
11.3.2.1.1. Chromohalobacter	0	0	0	0	0	0	0	8	0	0	0	0
11.3.2.1.2. Deleya	0	2	0	0	0	0	0	0	0	0	0	0
11.3.3. Pseudomonadales	300	21	70	0	47	2	72	439	80	483	44	317
11.3.3.1. Methylococcaceae	0	0	0	0	0	0	0	3	3	2	0	0
11.3.3.1.1. Methylomonas	0	0	0	0	0	0	0	3	3	2	0	0
11.3.3.2. Pseudomonadaceae	300	21	70	0	47	2	72	436	77	481	44	317
11.3.3.2.1. Azotobacter	0	0	0	0	0	0	0	82	1	1	4	13
11.3.3.2.2. Pseudomonas	300	21	70	0	47	2	72	354	76	480	40	304
11.3.4. Xanthomonadales	113	40	229	12	24	13	1	76	17	55	24	137
11.3.4.1. Xanthomonadaceae	113	40	229	12	24	13	1	76	17	55	24	137
11.3.4.1.1. Lysobacter	0	0	0	0	0	0	0	2	0	3	6	3
11.3.4.1.2. Pseudoxanthomonas	0	0	0	0	0	0	0	2	0	0	0	0
11.3.4.1.3. Rhodanobacter	0	0	0	0	0	0	0	2	0	0	0	0
11.3.4.1.4. Stenotrophomonas	0	0	0	0	0	0	1	0	0	0	0	0
11.3.4.1.5. Xanthomonas	113	40	229	12	24	13	0	72	17	52	18	134
11.4. Epsilonproteobacteria	4	0	0	0	1	0	1	2	0	1	0	0
11.4.1. Desulfobacterales	4	0	0	0	1	0	1	2	0	1	0	0
11.4.1.1. Geobacteraceae	4	0	0	0	1	0	0	0	0	0	0	0
11.4.1.1.1. Geobacter	4	0	0	0	1	0	0	0	0	0	0	0
11.4.1.2. Myxococcaceae	0	0	0	0	0	0	0	2	0	1	0	0
11.4.1.2.1. Anaeromyxobacter	0	0	0	0	0	0	0	2	0	1	0	0
11.4.1.2.2. Myxococcus	0	0	0	0	0	0	0	2	0	1	0	0
11.4.1.3. Sorangiaceae	0	0	9	0	0	0	1	0	0	0	0	0
11.4.1.3.1. Sporiangium	0	0	9	0	0	0	1	0	0	0	0	0
12. Spirochaetae	0	0	0	0	0	0	0	0	0	0	0	0
12.1. Spirochaetes	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1. Spirochaetales	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1.1. Spirochaetaceae	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1.1.1.Spirochaeta	0	0	0	0	0	0	0	0	0	0	0	0
13. Verrucomicrobia	0	0	0	0	0	0	0	0	0	0	0	0
13.1. Verrucomicrobia	0	0	0	0	0	0	0	0	0	0	0	0
13.1.1. Opitutae	0	0	0	0	0	0	0	0	0	0	0	0
13.1.1.1. Opitutaceae	0	0	0	0	0	0	0	0	0	0	0	0
13.1.1.1.1.Opitutus	0	0	0	0	0	0	0	0	0	0	0	0
13.2. Unclassified Verrucomicrobia	0	0	0	0	0	0	0	0	0	0	0	0
14. Unclassified Bacteria	1	22	60	6	129	21	315	2604	145	554	228	967

Table S2.2. Accession numbers in EBI and NCBI databases and taxonomy of reference sequences used for the primer design of the PHOD primers.

#	Accession number (EBI or NCBI)	Phylum Phylum	Order Order	Family Family	Species Species
1	ENA|EEI26313	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium glucuronolyticum* ATCC 51867
2	ENA|BAF55139	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium glutamicum*
3	ENA|BAB99658	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium glutamicum* ATCC 13032
4	ENA|AGF72794	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium halotolerans* DSM 44683
5	ENA|AGF73056	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium halotolerans* DSM 44683
6	ENA|AGF73652	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium halotolerans* DSM 44683
7	ENA|AGG67341	*Actinobacteria*	*Actinomycetales*	*Corynebacteriaceae*	*Corynebacterium halotolerans* DSM 44683
8	ENA|AFA75231	*Actinobacteria*	*Actinomycetales*	*Gordoniaceae*	*Gordonia polyisoprenivorans* VH2
9	ENA|GAC61596	*Actinobacteria*	*Actinomycetales*	*Gordoniaceae*	*Gordonia sihwensis* NBRC 108236
10	ENA|EMP12511	*Actinobacteria*	*Actinomycetales*	*Gordoniaceae*	*Gordonia* sp.
11	ENA|AGJ68675	*Actinobacteria*	*Actinomycetales*	*Mycobacteriaceae*	*Mycobacterium tuberculosis* NITR203
12	ENA|AFN64192	*Actinobacteria*	*Actinomycetales*	*Mycobacteriaceae*	*Mycobacteriumtuberculosiscomplex*
13	ENA|AFT99034	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Nocardia brasiliensis* ATCC 700358
14	ENA|AFU00244	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Nocardiabrasiliensis* ATCC 700358
15	ENA|AFU03643	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Nocardia brasiliensis* ATCC 700358
16	ENA|CCF63942	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Nocardia cyriacigeorgica*
17	ENA|EGD25772	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Rhodococcus equ* iATCC 33707
18	ENA|EJI95441	*Actinobacteria*	*Actinomycetales*	*Nocardiaceae*	*Rhodococcus* sp. JVH1
19	ENA|ABD10394	*Actinobacteria*	*Actinomycetales*	*Frankiaceae*	*Frankia* sp. ccI3
20	ENA|ABW10760	*Actinobacteria*	*Actinomycetales*	*Frankiaceae*	*Frankia* sp. EAN1
21	ENA|ADD44130	*Actinobacteria*	*Actinomycetales*	*Glycomycetaceae*	*Stackebrandtia nassauensis* DSM 44728
22	ENA|ADD45345	*Actinobacteria*	*Actinomycetales*	*Glycomycetaceae*	*Stackebrandtia nassauensis* DSM 44728
23	ENA|ACL40996	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter chlorophenolicus* A6
24	ENA|ACV06763	*Actinobacteria*	*Actinomycetales*	*Dermacoccaceae*	*Kytococcus sedentarius* DSM 20547
25	ENA|EAQ00942	*Actinobacteria*	*Actinomycetales*	*Intrasp.orangiaceae*	*Janibacter* sp. HTCC2649
26	ENA|CBT74545	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter arilaitenss*
27	ENA|EMR00121	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter gangotriensis*
28	ENA|AFR30629	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter* sp.
29	ENA|ABK02451	*Actinobacteria*	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter* sp. FB24
30	ENA|AEV83725	*Actinobacteria*	*Actinomycetales*	*Micromonosporaceae*	*Actinoplanes* sp. SE50
31	ENA|AEV87274	*Actinobacteria*	*Actinomycetales*	*Micromonosporaceae*	*Actinoplanes* sp. SE50
32	ENA|ADL48407	*Actinobacteria*	*Actinomycetales*	*Micromonosporaceae*	*Micromonospora aurantiaca* ATCC 27029
33	ENA|CCH17164	*Actinobacteria*	*Actinomycetales*	*Micromonosporaceae*	*Micromonospora lupini*
34	ENA|EEP71055	*Actinobacteria*	*Actinomycetales*	*Micromonosporaceae*	*Micromonospora* sp. ATCC 39149
35	ENA|ADB32048	*Actinobacteria*	*Actinomycetales*	*Nocardioidaceae*	*Kribbella flavida* DSM 17836
36	ENA|AEK40043	*Actinobacteria*	*Actinomycetales*	*Pseudonocardiaceae*	*Amycolatopsis mediterranei* S699
37	ENA|ADJ48441	*Actinobacteria*	*Actinomycetales*	*Pseudonocardiaceae*	*Amycolatopsis mediterranei* U32
38	ENA|BAC74775	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces avermitilis* MA-4680
39	ENA|ADI10639	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces bingchenggensis* BCW
40	ENA|ADI10779	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces bingchenggensis* BCW
41	ENA|ADI11949	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces bingchenggensis* BCW
42	ENA|ADI12233	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces bingchenggensis* BCW
43	ENA|EFG06649	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces clavuligerus* ATCC 27064
44	ENA|CAD55169	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces coelicolor*
45	ENA|CCK31695	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces davawensis* JCM4913
46	ENA|ADW06950	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces flavogriseus* ATCC 33331
47	ENA|BAG23089	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces griseus* NBRC13350
48	ENA|AEY87979	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces hygroscopicus*
49	ENA|CBG74762	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces scabiei*
50	ENA|AEN08616	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces* sp.
51	ENA|ELP68656	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces turgidiscabies*
52	ENA|CCA54151	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces venezuelae* ATCC 10712
53	ENA|AEM83182	*Actinobacteria*	*Actinomycetales*	*Streptomycetaceae*	*Streptomyces violaceus*
54	ENA|AFR06971	*Actinobacteria*	*Actinomycetales*	*Nocardiopsaceae*	*Nocardiopsis alba* ATCC BAA
55	ENA|AFR07079	*Actinobacteria*	*Actinomycetales*	*Nocardiopsaceae*	*Nocardiopsis alba* ATCC BAA
56	ENA|ADH67544	*Actinobacteria*	*Actinomycetales*	*Nocardiopsaceae*	*Nocardiopsis dassonvillei* DSM 43111
57	ENA|ACY99758	*Actinobacteria*	*Actinomycetales*	*Thermomonosporaceae*	*Thermomonosp.ora curvata* DSM 43183

Table S2.2. Accession numbers in EBI and NCBI databases and taxonomy of reference sequences used for the primer design of the PHOD primers. (Continued)

#	Accession number (EBI or NCBI)	Phylum	Order	Family	Species
58	ENA|ABG03284	*Actinobacteria*	*Rubrobacterales*	*Rubrobacteraceae*	*Rubrobacter xylanophilus* DSM 9941
59	ENA|ADB49402	*Actinobacteria*	*Solirubrobacterales*	*Conexibacteraceae*	*Conexibacter woesei* DSM 14684
60	ENA|EKB49993	*Bacteroidetes*	*Cytophagales*	*Cyclobacteriaceae*	*Cecembia lonarensis* LW9
61	ENA|AEL28382	*Bacteroidetes*	*Cytophagales*	*Cyclobacteriaceae*	*Cyclobacterium marinum* DSM 745
62	ENA|AEL28385	*Bacteroidetes*	*Cytophagales*	*Cyclobacteriaceae*	*Cyclobacterium marinum* DSM 745
63	ENA|AEL28386	*Bacteroidetes*	*Cytophagales*	*Cyclobacteriaceae*	*Cyclobacterium marinum* DSM 745
64	ENA|AFK04199	*Bacteroidetes*	*Cytophagales*	*Cytophagaceae*	*Emticicia oligotrophica* DSM 17448
65	ENA|CCH53780	*Bacteroidetes*	*Cytophagales*	*Cytophagaceae*	*Fibrisoma limi* BUZ3
66	ENA|ADR23105	*Bacteroidetes*	*Cytophagales*	*Flammeovirgaceae*	*Marivirga tractuosa* DSM 4126
67	ENA|AEA43205	*Bacteroidetes*	*Flavobacteriales*	*Cryomorphaceae*	*Fluviicola taffensis* DSM 16823
68	ENA|ADV50039	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Cellulophaga algicola* DSM 14237
69	ENA|ADV51026	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Cellulophaga algicola* DSM 14237
70	ENA|ADY28100	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Cellulophaga lytica* DSM 7489
71	ENA|EFK37022	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Chryseobacterium gleum* ATCC 35910
72	ENA|AEW86730	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacterium columnare* ATCC 49512
73	ENA|AEE19783	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Krokinobacter* sp. 4H-3-7-5
74	ENA|AEM69324	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Muricauda ruestringensis* DSM 13258
75	ENA|EAQ41320	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Polaribacter* sp. MED152
76	ENA|ADQ82175	*Bacteroidetes*	*Flavobacteriales*	*Flavobacteriaceae*	*Riemerellaanati pestifer* ATCC 11845
77	ENA|CCF99636	*Bacteroidetes*	*Flavobacteriia*	*Flavobacteriia*	*Flavobacteriia bacterium*
78	ENA|CBH25587	*Bacteroidetes*	*Incertaesedis(II)*	*Rhodothermaceae*	*Salinibacter ruber* M8
79	ENA|AEE48088	*Bacteroidetes*	*Sphingobacteriales*	*Saprosp.iraceae*	*Haliscomenobacte rhydrossis* DSM 1100
80	ENA|AEE49081	*Bacteroidetes*	*Sphingobacteriales*	*Saprosp.iraceae*	*Haliscomenobacte rhydrossis* DSM 1100
81	ENA|AFC23630	*Bacteroidetes*	*Sphingobacteriales*	*Saprospiraceae*	*Saprospira grandis*
82	ENA|AFZ31505	*Cyanobacteria*	*Chroococcales*	*Gloeocapsa*	*Gloeocapsa* sp. PCC7428
83	ENA|AFZ33354	*Cyanobacteria*	*Chroococcales*	*Gloeocapsa*	*Gloeocapsa* sp. PCC7428
84	ENA|AFZ44148	*Cyanobacteria*	*Chroococcales*	*Halothececluster*	*Halothece* sp. PCC7418A
85	ENA|ABD00626	*Cyanobacteria*	*Chroococcales*	*Synechococcus*	*Synechococcus* sp. JA-3-3Ab
86	ENA|ABD01669	*Cyanobacteria*	*Chroococcales*	*Synechococcus*	*Synechococcus* sp. JA-3-3Ab
87	ENA|AFZ57827	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Anabaena cylindrica* PCC7122
88	ENA|ABA21870	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Anabaena variabilis* ATCC 29413
89	ENA|ABA22311	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Anabaena variabilis*
90	ENA|ABA23729	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Anabaena variabilis*
91	ENA|ACC82149	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc punctiforme* PCC73102
92	ENA|AFY41332	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7107
93	ENA|AFY43518	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7107
94	ENA|BA000019	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7120
95	ENA|BAB76675	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7120
96	ENA|BAB77731	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7120
97	ENA|AFY49531	*Cyanobacteria*	*Nostocales*	*Nostocaceae*	*Nostoc* sp. PCC7524
98	ENA|AFZ03230	*Cyanobacteria*	*Nostocales*	*Rivulariaceae*	*Calothrix* sp. PCC6303
99	ENA|AFY52909	*Cyanobacteria*	*Nostocales*	*Rivulariaceae*	*Rivularia* sp. PCC7116
100	ENA|AFZ20635	*Cyanobacteria*	*Oscillatoriales*	*Microcoleus*	*Microcoleus* sp. PCC7113
101	ENA|ACK70604	*Cyanobacteria*	*Oscillatoriophycideae*	*Chroococcales*	*Cyanothece* sp. PCC-7424
102	ENA|AFY86481	*Cyanobacteria*	*Pleurocapsales*	*Chroococcidiopsis*	*Chroococcidiopsis thermalis* PCC7203
103	ENA|AFY90141	*Cyanobacteria*	*Pleurocapsales*	*Chroococcidiopsis*	*Chroococcidiopsis thermalis* PCC7203
104	ENA|ADY27657	*Deinococcus-Thermus*	*Deinococcales*	*Deinococcaceae*	*Deinococcus proteolyticus*
105	ENA|ADH62171	*Deinococcus-Thermus*	*Thermales*	*Thermaceae*	*Meiothermus silvanus* DSM 9946
106	ENA|AFZ89285	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus amyloliquefaciens subsp.Plantarum* AS43
107	ENA|AEB22432	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus amyloliquefaciens* TA208
108	ENA|ADP34908	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus atrophaeus* 1942
109	ENA|AAU39240	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus licheniformis* DSM 13
110	ENA|ADE72073	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus megaterium* QMB1551
111	ENA|AEN87067	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus megaterium* WSH-002
112	ENA|AFI26806	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus* sp. JS
113	ENA|EFV74755	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus* sp.. 2A57CT2
114	ENA|ADV95192	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus subtilis* BSn5
115	ENA|ADM36326	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus subtilis* subsp.sp.izizenii

Table S2.2. Accession numbers in EBI and NCBI databases and taxonomy of reference sequences used for the primer design of the PHOD primers. (Continued)

#	Accession number (EBI or NCBI)	Phylum	Order	Family	Species
		Phylum	Order	Family	Species
116	ENA|AEP85178	*Firmicutes*	*Bacillales*	*Bacillaceae*	*Bacillus subtilis* subsp.sp.izizenii TU-B-10
117	ENA|AFH59845	*Firmicutes*	*Bacillales*	*Paenibacillaceae*	*Paenibacillus mucilaginosus* K02
118	ENA|EGK13909	*Firmicutes*	*Bacillales*	*Thermoactinomycetaceae*	*Desmosp.ora* sp. 8437
119	ENA|BAH40216	*Gemmatimonadetes*	*Gemmatimonadales*	*Gemmatimonadaceae*	*Gemmatimonas aurantiaca*
120	ENA|AFH49936	*Ignavibacteriae*	*Ignavibacteriales*	*Ignavibacteriaceae*	*Ignavibacterium album* JCM16511
121	ENA|ADV63779	*Planctomycetes*	*Planctomycetales*	*Planctomycetaceae*	*Isosp.haera pallida* ATCC 43644
122	ENA|ADY59064	*Planctomycetes*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces brasiliensis* DSM 5305
123	ENA|ADY61457	*Planctomycetes*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces brasiliensis* DSM 5305
124	ENA|ADY61880	*Planctomycetes*	*Planctomycetales*	*Planctomycetaceae*	*Planctomyces brasiliensis* DSM 5305
125	ENA|AFS38873	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas macleodi* iATCC 27126
126	ENA|AEA99704	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas macleodii*
127	ENA|AFT76084	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas macleodii*
128	ENA|AFT79860	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas macleodii*
129	ENA|AEF05323	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Alteromonas* sp. SN2
130	ENA|GAC17848	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola arctica* BSs20135
131	ENA|GAC21094	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola arctica* BSs20135
132	ENA|GAC22211	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola arctica* BSs20135
133	ENA|GAC26248	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola mesophila* KMM241
134	ENA|GAC26997	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola pallidula* DSM 14239
135	ENA|GAC29575	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola pallidula* DSM 14239
136	ENA|GAC30922	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola pallidula* DSM 14239
137	ENA|GAC33715	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola polaris* LMG21857
138	ENA|GAC38437	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola psychrophila* 170
139	ENA|GAB56614	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Glaciecola punicea* DSM 14233
140	ENA|CCG96968	*Proteobacteria*	*Alteromonadales*	*Alteromonadaceae*	*Marinobacter hydrocarbonoclasticus* ATCC 49840
141	ENA|ADN76004	*Proteobacteria*	*Alteromonadales*	*Ferrimonadaceae*	*Ferrimonas balearica* DSM 9799
142	ENA|GAA62951	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20311
143	ENA|GAA66449	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20429
144	ENA|GAA69043	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20429
145	ENA|GAA73690	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20480
146	ENA|GAA80269	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20495
147	ENA|GAA81571	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20495
148	ENA|GAA61442	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. BSi20652
149	ENA|ADT68141	*Proteobacteria*	*Alteromonadales*	*Pseudoalteromonadaceae*	*Pseudoalteromonas* sp. SM9913
150	ENA|ABS09779	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella baltica* OS185
151	ENA|ABI73423	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella frigidimarina*
152	ENA|ABZ75449	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella halifaxensis* HAW-EB4
153	ENA|ABV86147	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella pealeana* ATCC 700345
154	ENA|ABO22704	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanellaloihica* pv.-4
155	ENA|ACA85252	*Proteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanellawoodyi ATCC* ATCC 51908
156	ENA|AET91953	*Proteobacteria*	*Burkholderiales*	*Burkholderiaceae*	*Burkholderia* sp. YI23
157	ENA|ABF09460	*Proteobacteria*	*Burkholderiales*	*Burkholderiaceae*	*Cupriavidus metallidurans* CH3
158	ENA|AEI79776	*Proteobacteria*	*Burkholderiales*	*Burkholderiaceae*	*Cupriavidus necator* N-1
159	ENA|AFU45370	*Proteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Acidovorax* sp. KKS102
160	ENA|ACY34046	*Proteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Comamonas testosteroni*
161	ENA|ACY34739	*Proteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Comamonas testosteroni*
162	ENA|ABX34995	*Proteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Delftia acidovorans* sp.H-1
163	ENA|ABM36524	*Proteobacteria*	*Burkholderiales*	*Comamonadaceae*	*Polaromonas naphthalenivorans* CJ2
164	ENA|AEK63472	*Proteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	*Collimonas fungivorans* Ter331
165	ENA|AEK63890	*Proteobacteria*	*Burkholderiales*	*Oxalobacteraceae*	*Collimonas fungivorans* Ter331
166	ENA|EDX79900	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Brevundimonas* sp. BAL3
167	ENA|EDX80389	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Brevundimonas* sp. BAL3
168	ENA|EDX81672	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Brevundimonas* sp. BAL3
169	ENA|AAK22442	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Caulobacter crescentus* CB15
170	ENA|AAK23544	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Caulobacter crescentus* CB15
171	ENA|ACG78853	*Proteobacteria*	*Caulobacterales*	*Caulobacteraceae*	*Phenylobacterium zucineum* HLK1
172	ENA|EGM78671	*Proteobacteria*	*Chromatiales*	*Chromatiaceae*	*Rheinheimera* sp. A13
173	ENA|EED33300	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria*	*gammaproteobacterium NOR5*

Table S2.2. Accession numbers in EBI and NCBI databases and taxonomy of reference sequences used for the primer design of the PHOD primers. (Continued)

#	Accession number (EBI or NCBI)	Phylum	Order	Family	Species
		Phylum	Order	Family	Species
174	ENA|EED35601	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria*	*gammaproteobacterium NOR51*
175	ENA|EKO36824	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria*	*SAR86clusterbacterium*
176	ENA|EKO37313	*Proteobacteria*	*Gammaproteobacteria*	*Gammaproteobacteria*	*SAR86clusterbacterium*
177	ENA|AEG01366	*Proteobacteria*	*Methylococcales*	*Methylococcaceae*	*Methylomonas methanica* MC09
178	ENA|AEI62293	*Proteobacteria*	*Myxococcales*	*Myxococcaceae*	*Myxococcus fulvus* HW-1
179	ENA|AGC45053	*Proteobacteria*	*Myxococcales*	*Myxococcaceae*	*Myxococcus stipitatus* DSM 14675
180	ENA|AAQ61942	*Proteobacteria*	*Neisseriales*	*Neisseriaceae*	*Chromobacterium violaceum* ATCC 12472
181	ENA|EDX89406	*Proteobacteria*	*Oceanosp.irillales*	*Alcanivoracaceae*	*Alcanivorax* sp. DG881
182	ENA|EDX91210	*Proteobacteria*	*Oceanosp.irillales*	*Alcanivoracaceae*	*Alcanivorax* sp. DG881
183	ENA|ABC28066	*Proteobacteria*	*Oceanosp.irillales*	*Hahellaceae*	*Hahella chejuensis* KCTC2396
184	ENA|ABC31544	*Proteobacteria*	*Oceanosp.irillales*	*Hahellaceae*	*Hahella chejuensis* KCTC2396
185	ENA|ABE60068	*Proteobacteria*	*Oceanosp.irillales*	*Halomonadaceae*	*Chromohalobacter salexigens* DSM 3043
186	ENA|CBV42268	*Proteobacteria*	*Oceanosp.irillales*	*Halomonadaceae*	*Halomonas elongata* DSM 2581
187	ENA|EEB78901	*Proteobacteria*	*OMGgroup*	*OM60clade*	*Gammaproteobacteria*
188	ENA|ACJ56861	*Proteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter baumannii* AB307-0294
189	ENA|EKU56649	*Proteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter baumannii* WC-323
190	ENA|ADI89668	*Proteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter oleivorans* DR1
191	ENA|ELW83481	*Proteobacteria*	*Pseudomonadales*	*Moraxellaceae*	*Acinetobacter* sp. WC-743
192	ENA|ACO77444	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Azotobacte rvinelandii*
193	ENA|ABR86870	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas aeruginosa* Pa7
194	ENA|ABJ13183	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas aeruginosa* UCBPP-PA14
195	ENA|EKG33501	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas avellanae* BPIC631
196	ENA|AEA66958	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas chlororaphis*
197	ENA|EIM16116	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas chlororaphis* O6
198	ENA|EJL08915	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas chlororaphissubsp*
199	ENA|AFJ58882	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas fluorescens* A506
200	ENA|AFJ59388	*Proteobacteria*	*Pseudomonadales*	*Pseudomonadaceae*	*Pseudomonas fluorescens* A506

3

PHOX ALKALINE PHOSPHATASE GENE DIVERSITY IN SOIL

Combined with Chapter 4 and submitted as:

Ragot SA, MA Kertesz, É Mészáros, E Frossard and EK Bünemann. Soil *phoD* and *phoX* alkaline phosphatase gene compositions respond to multiple environmental factors. *FEMS Microbiology Ecology*.

Abstract

The soil microbial community contributes to phosphorus cycling, e.g. by secreting PhoX phosphatases. PhoX is a monomeric enzyme that principally hydrolyzes phosphomonoesters, which is the main type of soil organic P. To date, however, our knowledge on the composition and structure of the *phoX*-harboring community in the environment is poor, due to the lack of molecular tools. Here, we assessed our current knowledge of the taxonomic and environmental distribution of *phoX* genes in the environment using whole-genome and metagenome databases, and we designed new *phoX*-targeting primers to study the *phoX* gene in soil.

The primer set was validated by 454-sequencing of six soils collected from two continents with different climates and soil properties. In the databases, the *phoX* gene was reported in 15 bacterial phyla and was most abundant in metagenomes from terrestrial ecosystems, followed by marine ecosystems and microbiomes associated with plants. The newly-designed primers amplified *phoX* from microorganisms affiliated to 1 archaeal and 16 bacterial phyla. *phoX*-harboring *Actinobacteria* and *Proteobacteria* were found in high abundances in all soil samples. However, *phoX*-harboring *Acidobacteria, Actinobacteria, Chloroflexi, Planctomycetes, Proteobacteria* and *Verrucomicrobia* were also dominant in certain soils, resulting in contrasting *phoX*-harboring community compositions between samples. A two-fold difference in *phoX* species richness was found between the samples, with up to 214 *phoX* operational taxonomic units per sample. Redundancy analysis indicated that soil pH and total P were significantly correlated with the *phoX*-harboring community structure.

Our results show that the newly-designed primers are a valuable tool to study the *phoX*-harboring community in environmental samples.

3.1 Introduction

The soil microbial community contributes to phosphorus (P) cycling by secreting phosphatases. Phosphatases catalyze the hydrolysis of organic P and can hydrolyze up to 89% of the total extracted organic P (Jarosch et al. 2015), with gross organic P mineralization rates ranging between 0.1 and 2.5 mg P $kg^{-1}d^{-1}$ and up to 12.6 mg P $kg^{-1}d^{-1}$ (Bünemann 2015). Hence, they play a role in the replenishment of orthophosphate in the soil solution, which is the only available P form for plants and microorganisms in soil. To date, three alkaline phosphatase families, PhoA, PhoD and PhoX, have been described and are characterized by the hydrolysis of phosphomonoesters, which is generally the dominant type of organic P in soil (Condron et al. 2005). PhoX is a monomeric enzyme commonly activated by two Fe^{2+} and three Ca^{2+} as co-factors with an optimum enzymatic activity at pH 7.5 to 10 (Van Mourik et al. 2008; Kathuria and Martiny 2011; Liao and Siegbahn 2015). Although PhoX is principally described as a phosphomonoesterase, the substrate affinity of PhoX can vary depending on the species. For example, the PhoX isolated from *Pseudomonas multocida* X-73 can hydrolyze both phosphomono- and phosphodiesters (Wu et al. 2007), whereas the PhoX produced by *Sinorhizobium meliloti* can hydrolyze phosphomonoesters and N-P bonds (Sebastián and Ammerman 2011; Zaheer et al. 2009). PhoX can be found in the cytoplasm, in the periplasm, anchored in the outer-membrane or free in the environment (Luo et al. 2009). It has been reported to be secreted via the TAT secretion pathway in some bacteria (Wu et al. 2007).

The presence of *phoX* in aquatic ecosystems including fresh water and marine ecosystems has been reported by many studies (Luo et al. 2009; Sebastián and Ammerman 2009; Dai et al. 2014), In marine bacteria, the *phoX* gene is most frequently found in *Proteobacteria*, mainly *Alpha-* and *Gamma-Proteobacteria*, and in *Cyanobacteria*, but has also been found in some *Actinobacteria, Alteromonadales, Bacteroidetes, Lentisphaerae* and *Planctomycetes* (Luo et al. 2009; Sebastián and Ammerman 2009). The *phoX* gene has been mainly reported in bacteria, but has also been found in the eukaryotic microor-

ganisms *Chlamydomonas reinhardtii* (Moseley et al. 2006) and *Volvox cateri* (Hallmann 1999).

The first *phoX*-targeting primers were designed by Sebastián and Ammerman (2009) based on sequences of marine bacteria such as *Vibrio* sp., *Shewanella* sp. and *Roseobacter* sp. Using qPCR, they showed that *phoX* is commonly found in the Sargasso Sea and the Chesapeake Bay. In Lake Taihu (China), the abundance of *phoX* genes has been reported to be 2-fold higher in mesotrophic than eutrophic regions, while the diversity of *phoX* was 1.2-fold higher in eutrophic than in mesotrophic region (Dai et al. 2014). Besides aquatic ecosystems, the *phoX* gene has also been reported in soil bacteria such as *Sinorhizobium meliloti* (Zaheer et al. 2009). Nevertheless, our knowledge on the *phoX* gene in ecosystems other than aquatic ones is very poor, mainly due to the lack of molecular tools.

Here, we assessed our current knowledge on the environmental prevalence and taxonomic distribution of the *phoX* gene using the Integrated Microbial Genomes/Metagenomes (IMG/M) database. Additionally, we designed a new set of primers that target *phoX* in soil microorganisms and tested them on six soils collected on two continents.

3.2 Material and Methods

3.2.1 Taxonomic and environmental prevalence of *phoX* alkaline phosphatase genes – a meta-analysis

The prevalence of the *phoX* gene was assessed using the Integrated Microbial Genomes and Metagenomes (IMG/M) database, a dedicated system for annotation of whole genomes and metagenomes (Markowitz et al. 2012). Draft and complete genome datasets were used to evaluate the distribution of *phoX* across microbial phyla, and metagenome datasets were used to evaluate the prevalence of *phoX* in the environment (data accessed on September 10th 2015). Metagenome datasets were normalized as number of *phoX*

counts per number of bases sequenced per metagenome dataset and categorized after the classification of the environmental types in the IMG/M database.

3.2.2 Soil sampling and general soil properties

Six grassland soils (S1-6) were collected from different locations in Australia in spring 2013 (S1-4) and in Switzerland in summer 2014 (S5-6) (Table 3.1), covering a broad range of soil groups and climatic conditions. For each sampling, five soil cores from the top 5 cm were randomly collected and homogenized by sieving (4 mm). Subsamples were stored at -20°C for molecular analysis. The remaining composite soil was air dried and used to determine the basic soil properties, including pH, potential alkaline phosphatase activity, total organic carbon (TOC) and total P (TP). Soil pH was measured in a soil suspension in 0.01 M $CaCl_2$ with a 1:2.5 mass/volume ratio using a Benchtop pH/ISE 720A (Orion Research Inc., Jacksonville, FL). TOC was measured on dry and ground soil using a CNS analyzer (Thermo-Finnigan). No carbonate was detected in the soils. TP was determined by wet digestion with H_2O_2-H_2SO_4 (Anderson and Ingram 1993) and measured with malachite green at 610 nm (Ohno and Zibilske 1991). Potential alkaline phosphatase activity was determined following Marx et al. (2001) as modified by Poll et al. (2006). Soil pH ranged between 4.2 and 6.8. TOC varied between 5 and 36 g kg^{-1} soil, and TP varied between 193 and 705 mg kg^{-1} soil. Potential alkaline phosphatase activity varied between 22 and 3,668 nmol substrate g^{-1} soil h^{-1}.

3.2.3 DNA extraction from soil

Genomic DNA was extracted from 0.25 g frozen soil using PowerSoil®DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) following the manufacturer's instructions with a slight modification in the homogenization and cell lysis step: TissueLyzer II (Qiagen, Valencia, CA, USA) was used twice for 3 min at 30 Hz. DNA

Table 3.1. Location and soil group of the sampled soils.

#	Sample name	Location	Geographical coordinates[1]	Soil group[2]	pH	TOC	TP	Potential alkaline phosphatase activity
						g kg^{-1} soil	mg kg^{-1}	nmol substrate g^{-1} soil h^{-1}
1	S1	Kia-Ora, Australia	34 ° 48’ 18"S, 148 ° 35’ 0"E	Planosol	4.2±0.3	21.0±0.8	211±8	43±3
2	S2	Narrabi, Australia	30 ° 15’14”S, 149 ° 51’53”E	Planosol	6.2±0.0	23.7±0.1	705±13	720±58
3	S3	Nyngan, Australia	31 ° 25’52"S, 147 ° 4’9"E	Cambisol	4.7±0.1	15.0±0.3	466±10	22±4
4	S4	Mutawintji, Australia	31 ° 16’19"S, 142 ° 17’44"E	Leptosol	6.8±0.1	5.0±0.2	193±11	177±29
5	S5	Watt, Switzerland	47 ° 25’45”N, 8 ° 29’31”E	Cambisol	6.1±0.1	36.1±0.1	700±99	3,668±148
6	S6	Reckenholz, Switzerland	47 ° 25’36”N, 8 ° 31’9”E	Luvisol	5.2±0.1	22.4±0.4	457±32	436±27

[1]Köppen-Geiger climate classification. Climate categories are described further in a paper by Kottek et al. (2006).
[2]World Reference Base for Soil Resource (IUSS 2014).

concentration and quality were assessed with a micro-volume UV/VIS spectrophotometer (UVS-99, UVISDROP; ACTGene, Inc., Piscataway, NJ, USA).

3.2.4 Design and validation of the *phoX* alkaline phosphatase-targeting primers

Gene sequences annotated as *phoX* and/or associated with COG3211 (Cluster of Orthologous Groups; http://ww.ncbi.nlm.nih.gov/COG/), which corresponds to *phoX* alkaline phosphatase, were retrieved from the databases European Nucleotide Archive (ENA) and Uniprot Knowledgebase (UniprotKB). They were then clustered at 97% similarity using CD-HIT (Huang et al. 2010), resulting in a total of 228 sequences used as reference for the primer design (see list in Table S3.1 and taxonomic tree in Figure S3.1). Sequences were affiliated to 8 phyla including *Actinobacteria* (48 sequences), *Aquificae* (1 sequence), *Bacteroidetes* (1 sequence), *Chloroflexi* (6 sequences), *Cyanobacteria* (27 sequences), *Deinococcus-Thermus* (1 sequence), *Firmicutes* (1 sequence) and *Proteobacteria* (143 sequences).

Alignment was constructed using MUSCLE (Edgar 2004) and manually reviewed and improved in Geneious 6.1.2 (Biomatters, Australia, http://www.geneious.com). The most suitable regions for primer design were identified using PrimerProspector (Walters et al. 2011). Forward and reverse candidate primers were then manually designed to reach the maximum coverage of the reference sequences. Candidate primer pairs were tested in-silico using De-metaST-BLAST (Gulvik et al. 2012) to identify potential primer pairs with an appropriate product size and coverage of the reference sequences.

Candidate primers (11 forward and 13 reverse primers) were tested in a gradient PCR using a mixture of soil genomic DNA containing soil from S5-6 as template (Table 3.1). PCR reactions were carried out in a 25 µl volume containing 1X MyTaq Reaction Buffer (including $MgCl_2$ and dNTPs), 0.5 µM of each primer and 0.6 Units of MyTaq polymerase (Bioline, NSW, Australia) with 1-2 ng DNA as template in an S1000 thermocycler (Bio-Rad Laboratories, CA). The amplification reaction included a gradient thermal profile

with an annealing temperature of ± 3 degrees from the optimal annealing temperature calculated for each candidate primer pair. Amplicon size and intensity, and the presence of primer dimers, was assessed visually by electrophoresis on a 1.5% (w/v) agarose gel stained with ethidium bromide.

Amplicon coverage and specificity was evaluated for the most successful primer pairs by cloning and sequencing. PCR products were ligated at 4 °C overnight using pGEM®-T Vector Systems (Promega, Madison, WI) and transformed into chemically competent *E. coli* cells (α-select; F-*(*deo)R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17 (rk^-, mk^+) E44 thi-1 *pho*A Δ(*lac*ZYA-*arg*F) U169 Φ80*lac*Z ΔM15 λ^-) following the manufacturer's instructions (Bioline, NSW, Australia). Restriction fragment length polymorphism (RFLP) profiling of clones with the expected insert size was done using *Hha*I (0.2 U μl^{-1} for 3 hours at 37°C) (Promega), and profiles were visualized by electrophoresis on a 2% agarose gel. Representative inserts of unique RFLP profiles were then sequenced (Macrogen Inc., Seoul, Korea). The resulting sequences were used to evaluate the coverage and specificity of the candidate primer pairs using BLAST search (Altschul et al. 1990). Finally, the primer set *phoX*-F455 (5'- CAG TTC GGB TWC AAC AAC GA-3')/*phoX*-R1076 (5'- CGG CCC AGS GCR GTG YGY TT -3'), named after the position in the *phoX* gene in *Rhodococcus opacus* B4, had the highest diversity and coverage.

Additionally, the coverage of the primer set *phoX*-F455/R1076 was compared to the three *phoX*-targeting primer sets (*phoX*1, *phoX*2, *phoX*3) published by Sebastián and Ammerman (2009). The coverage of the *phoX* reference sequences used for this study by each forward and reverse primers and by each primer set was evaluated in-silico in Geneious 6.1.2. The primer set *phoX*-F455/R1076 had the highest coverage and targeted 2.5 times more sequences than the other primers (Table S3.3). *phoX*-F455/R1076 was selected for 454-sequencing on the six soil samples (Table 3.1).

Prior to sequencing, duplicate DNA extracts were tested for presence of *phoX* using PCR. PCR reactions were performed in a 25 µL volume containing 1X GoTaq Reaction Buffer (including $MgCl_2$ and dNTPs), 0.5 µM of each primer and 0.6 Units of GoTaq

polymerase (Promega) with 1-2 ng DNA as template in a Labcycler (Sensoquest, Göttingen, Germany). The amplification reaction included an initial denaturation step of 5 min at 95°C followed by 35 cycles of a denaturation step of 30 s at 95°C, an annealing step of 30 s at 60°C and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. The quality of the amplification was assessed visually by electrophoresis on a 1.5% (w/v) agarose gel stained with ethidium bromide.

Duplicate DNA extracts were then pooled and adjusted to a concentration of 20 ng μL^{-1}. Pooled extracts were sent for 454-sequencing on a GS-FLX+ platform (Roche 454 Life Sciences, Branford, CT) by Research and Testing Laboratory (Lubbock, TX, http://www.researchandtesting.com). Barcoded primer design, sequencing and initial quality filtering was performed by Research and Testing Laboratory using standard protocols. Briefly, sequences with a quality score below 25 were trimmed. Chimeras were removed using USEARCH with a clustering at a 4% divergence (Edgar et al. 2011). Denoising was performed using the Research and Testing Denoiser algorithm. It uses the non-chimeric sequences and the quality scores to create consensus clusters from aligned sequences. Among each cluster, the probability of prevalence of each nucleotide is calculated and accordingly, a new quality score generated, which is then used to remove noise from the dataset.

3.2.5 Sequence analysis

Sequencing datasets were analyzed using MOTHUR (Schloss et al. 2009). Sequences were analyzed as nucleic acid sequences to keep the maximum information and allow accurate identification. After demultiplexing, reads including ambiguities, mismatches with the specific primers and the barcode were removed. Reads with an average quality score below 20 were then filtered. Resulting reads were trimmed at 200 bp and 600 bp as minimum and maximum length.

Sequences were aligned using the Needleman-Wunsch global alignment algorithm as implemented in MOTHUR, using 6-mers searching and the aligned reference sequences as template. The pairwise distance matrix was calculated from the alignment. The similarity cutoff to define *phoX* OTUs was determined using the same approach as in Tan et al. (2013). Briefly, 20 *phoX* sequences, representing the diversity of the *phoX* gene, and the 16S RNA gene of their corresponding bacterial strains were aligned using pairwise alignment (Table S3.1, Figure S3.1). The similarity matrices of *phoX* and 16S rRNA gene alignments were extracted and plotted against each other (Figure S3.2). The similarity cutoff corresponding to 97% similarity of the 16S RNA gene was at 57% similarity for the *phoX* gene. OTU matrices were normalized to the smallest library size using the normalized.shared command in MOTHUR to allow comparison between samples, by dividing the relative abundance of each OTU by the total number of reads per sample. The normalized values were then rounded to the nearest integer.

Taxonomy assignment was performed using blastn in BLAST+ (Camacho et al. 2009) with a minimum e-value of 1e-8 to retrieve NCBI sequence identifiers (GI accession number). Subsequently, in-house Perl scripts were used to populate and query a mysql database containing the NCBI GI number and taxonomic lineage information (script written by Stefan Zoller, Genetic Diversity Centre, ETH Zurich, available on request) (Table S3.1).

3.2.6 Data analysis

Observed species richness (S_{obs}), estimated species richness based on a library size of 3,000 reads (S_{est}), and the Chao1 species richness index (Chao and Shen 2003) were calculated based on standardized data using rarefaction curves in EstimateS (version 9; http: //purl.oclc.org/estimates). Additionally, the Good's coverage (Good 1953) and the alpha diversity estimated by the Shannon-Wiener index (H') (Gotelli and Colwell 2011) were calculated.

Correlations between the community composition and environmental variables were tested by redundancy analysis (RDA), followed by an analysis of variance (ANOVA) on the RDA fit using the vegan package (vegan; Community Ecology Package, **R** package version 2.2-0) in **R** version 2.15.0 (**R** Core Team, 2014; http://CRAN.R-project.org). Prior to analysis, the measured soil properties (TOC, TP, pH, alkaline phosphatase activity) were standardized using the Z-score method.

3.3 Results and Discussion

3.3.1 Taxonomic distribution of the *phoX* alkaline phosphatase gene

Our current knowledge of the taxonomic distribution of the *phoX* gene was evaluated using the IMG/M database. The *phoX* gene was reported in 5,511 classified bacterial genomes, but not in archaeal and fungal genomes, which is in more archaeal and bacterial phyla than previously reported in the IMG/M database. The *phoX* gene was found in 15 bacterial phyla (Figure 3.1). Among the *phoX*-harboring *Proteobacteria, Gammaproteobacteria* were the most frequently reported in the IMG/M database, representing 69% of the *Proteobacteria*, followed by *Alpha-Proteobacteria* (16%), *Beta-Proteobacteria* (8%), *Epsilon-Proteobacteria* (3%) and *Delta-Proteobacteria* (1%). The number of copies per genome varied between one and five, with 88% containing only one *phoX* homologue copy. Of all genomes available in the database, more than half of the *Cyanobacteria, Spirochetes* and *Tectomicrobia* harbored at least one copy of a *phoX* homologue (Figure 3.1).

The *phoX* gene has been reported in *Actinobacteria, Bacteroidetes, Cyanobacteria, Lentisphaerae, Planctomycetes* and *Proteobacteria* in marine bacteria (Luo et al. 2009; Sebastián and Ammerman 2009). Based on a GenBank search, Zaheer et al. (2009) found *phoX* bacteria such as *Actinobacteria, Cyanobacteria, Chloroflexi* and *Proteobacteria*. In agreement with our meta-analysis, these studies report high frequency of the *phoX* gene

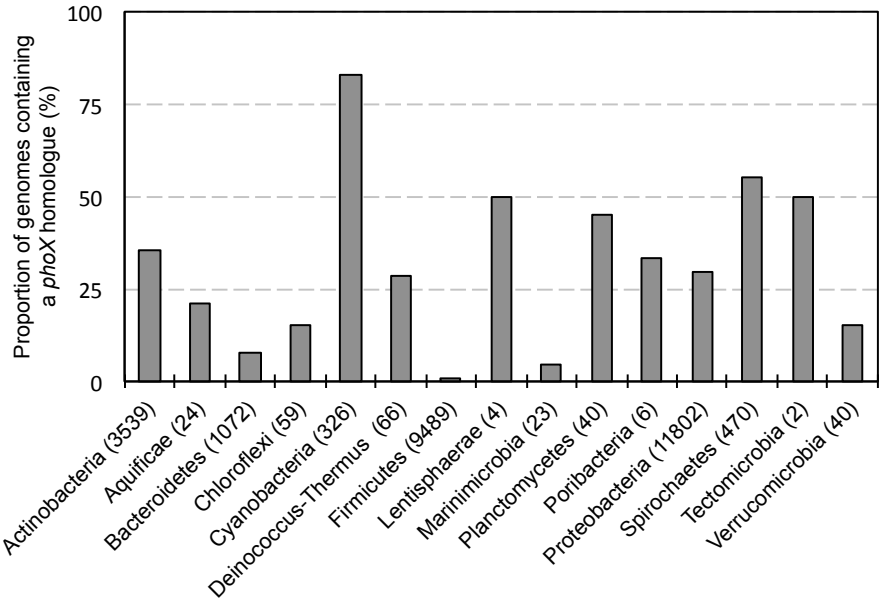


Figure 3.1. Proportion of sequenced genomes containing a least one *phoX* homologue in the IMG/M database (on September 10th 2015). Numbers in brackets indicate the total number of sequenced genomes in each phylum.

in *Actinobacteria* and *Proteobacteria*. Our results on the taxonomic distribution of the *phoX* gene suggest that *phoX* is more widespread across bacterial phyla than published studies indicate.

3.3.2 Environmental prevalence of the *phoX* alkaline phosphatase gene

The prevalence of the *phoX* gene was investigated in the 4,181 metagenomes available in the IMG/M database. The *phoX* gene was the most abundant in terrestrial ecosystems, followed by marine ecosystems and microbiomes associated with plants (Figure 3.2).

Previous metagenomic studies of water samples collected from the Northwest Atlantic to Eastern tropical Pacific oceans and North Pacific subtropical gyre showed that the *phoX* gene is frequently found in the ocean (Luo et al. 2009; Sebastián and Ammerman 2009).

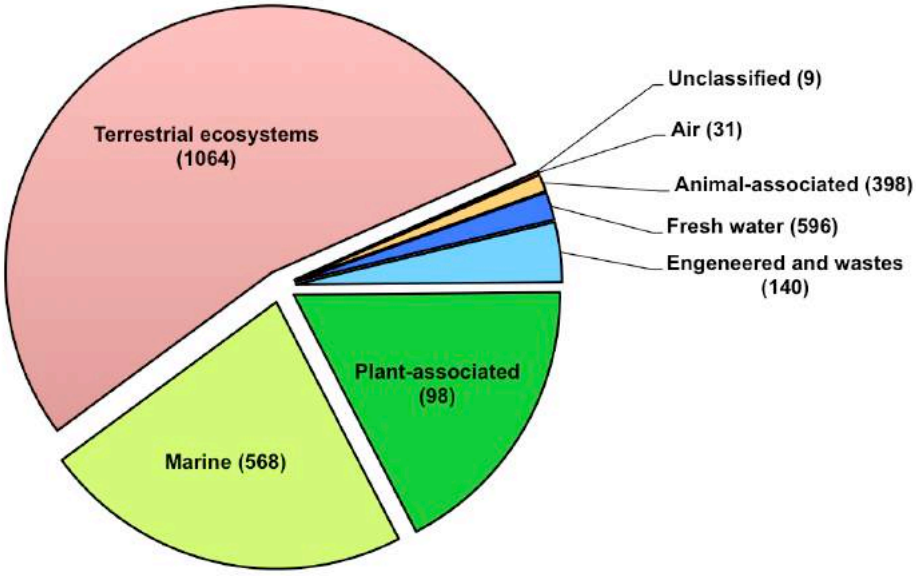


Figure 3.2. Relative abundance of *phoX* homologues in different types of environments (normalized as number of *phoX* counts per number of bases sequenced per metagenome dataset). Numbers in brackets indicate the number of metagenome datasets per environment type.

The *phoX* gene has also been reported in fresh water ecosystems under mesotrophic and eutrophic conditions (Dai et al. 2014). Although no metagenomic studies on *phoX* has yet been done in terrestrial ecosystems, the *phoX* gene has been found in common soil and rhizosphere bacteria like *Pseudomonas fluorescens, Pasteurella multocida, Sinorhizobium meliloti* and *Campylobacter jujeni* (Wu et al. 2007; Van Mourik et al. 2008; Zaheer et al. 2009), supporting the results of our meta-analysis.

3.3.3 Species richness, diversity and taxonomic composition of the *phoX*-harboring community in six grassland soils using newly designed primers

Species richness varied 2-fold between the samples (Table 3.2). The S_{obs} was lowest in S1 and highest in S6, with 47 and 98 OTUs, respectively. Chao1 and S_{est} indices showed a similar trend. Shannon diversity (H') was the lowest in S1 and the highest in S5. The amplified *phoX* genes were affiliated to 1 archaeal and 16 bacterial phyla (Figure 3.3, Table S3.2). Between 0 and 3.6% of the sequences in the samples could not be classified. Our results showed that the composition of the *phoX*-harboring communities varied greatly between samples (Figure 3.3, Table 3.2). For example, the *phoX*-harboring community in S2 was composed of 4 different phyla only, while that in S4 was composed of 14 different phyla. Nonetheless, *phoX*-harboring *Actinobacteria* and *Proteobacteria* were found in high abundances in all soils (Figure 3.3). In certain soils, *phoX*-harboring *Acidobacteria, Actinobacteria, Chloroflexi, Planctomycetes, Proteobacteria* and *Verrucomicrobia* were also dominant (Figure 3.3). Among the *phoX*-harboring *Proteobacteria, Alphaproteobacteria* were the most abundant class, followed by *Gamma-, Beta-* and *Deltaproteobacteria*, representing up to 99, 63 and 53% of the Proteobacteria in the samples (Table S3.2). S1 had a contrasting *phoX*-community composition compared to the other samples, with a rather high abundance of *phoX*-harboring *Acidobacteria* and no *phoX*-harboring *Planctomycetes*. S2 was characterized by a high abundance of *phoX*-harboring *Firmicutes*. S3 had the highest abundance of *phoX*-harboring *Chloroflexi* and the lowest abundance of *phoX*-harboring *Proteobacteria*, while S4 had the highest abundance of *phoX*-harboring *Cyanobacteria* and S5 the highest abundance of *phoX*-harboring *Bacteroidetes*. Finally, S6 was characterized by a high abundance of *phoX*-harboring *Armatimonadetes* and *Planctomycetes*.

Some of the dominant *phoX*-harboring groups found in our soil samples were also dominant in marine microbial communities such as *Actinobacteria, Cyanobacteria, Chloroflexi* and (*Alpha-, Beta- and Gamma-*) *Proteobacteria* (Sebastián and Ammerman 2009).

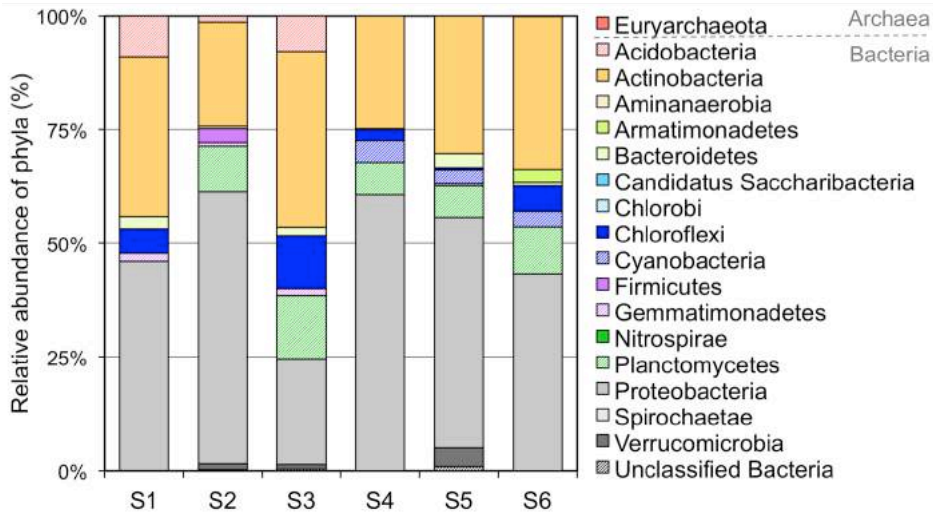


Figure 3.3. Relative abundance of *phoX*-harboring archaeal and bacterial phyla in 6 grassland soils.

In contrast to the results of the meta-analysis, the *phoX* gene was not amplified from *Tectomicrobia, Lentisphaerae* and *Marinimicrobia* in our soil samples. These differences in composition may be due to the environmental distribution of *phoX*-harboring microorganisms. For example, *Lentisphaerae*, a recently discovered phylum, are believed to inhabit mainly aquatic ecosystems (Cho et al. 2004), *Tectomicrobia* are primarily found in anaerobic terrestrial ecosystems and in microbiomes associated with marine sponges (Wilson et al. 2014), and *Marinamicrobia* in marine and fresh water environment (Coveley et al. 2015; Mueller et al. 2015). On the other hand, our primers amplified *phoX* from microorganisms affiliated to *Acidobacteria, Gemmatimonadetes* and *Candidatus-Saccharibacteria* that were not reported in aquatic microorganisms. Moreover, our primers amplified *phoX* from *Euryarchaeota* (Figure 3.3). Alkaline phosphatase activity has been reported in several archaea (Wende et al. 2010), however, to our knowledge, alkaline phosphatase in archaea has not yet been associated with the *phoX* gene. Additionally, the fact that *phoX* has been mostly amplified in bacteria may be attributed to the primer design, which was based on bacterial sequences only (Table S3.1). Our

Table 3.2. Species richness, Good's coverage, alpha diversity and taxonomy summary in the studied soil samples[1].

Soil sample	Species richness index			Good's	H'	Number of:			
	S_{obs}	S_{est}	Chao1	coverage		Phyla	Orders	Families	Genera
S1	47	117	102.2	0.88	3.1	14	16	24	29
S2	68	212	106	0.84	4.2	4	9	10	69
S3	96	109	205	0.62	3.7	11	11	16	39
S4	83	198	189	0.7	3.9	6	13	15	50
S5	90	131	129	0.75	4.3	5	17	22	68
S6	98	225	214	0.63	4.2	7	17	22	75

[1]Species richness indices (S_{obs}, S_{est}, and Chao1), Good's coverage, alpha diversity (Shannon-Wiener index, H'), and taxonomy (numbers of phyla, classes, orders, families, and genera).

results suggest that the *phoX*-harboring community in soil is rather diverse and is similar in composition to what has been previously reported in studies in aquatic ecosystems, suggesting that terrestrial and aquatic ecosystems harbor similar key *phoX*-harboring microorganisms.

3.3.4 Soil pH and total P are important determinants of the *phoX*-harboring community structure

Redundancy analysis (RDA) showed that S2, S3, S5 and S6 clustered together, whereas S1 and S4 clustered away from the other samples (Figure 3.4). Among the soil properties, soil pH and TP were significantly correlated with the *phoX*-harboring community structure. Soil pH was mainly associated with the first component, while TP was mainly associated to the second component. Together the first and second component explained 70% of the variation between the samples (Figure 3.4). Both S1 and S4 had low TP content, but different pH, which may explain their distinct distribution in the RDA (Table 3.1). Our results suggest that soil pH and TP strongly impact the *phoX*-harboring community structure, but this needs to be verified on a larger range of soil samples.

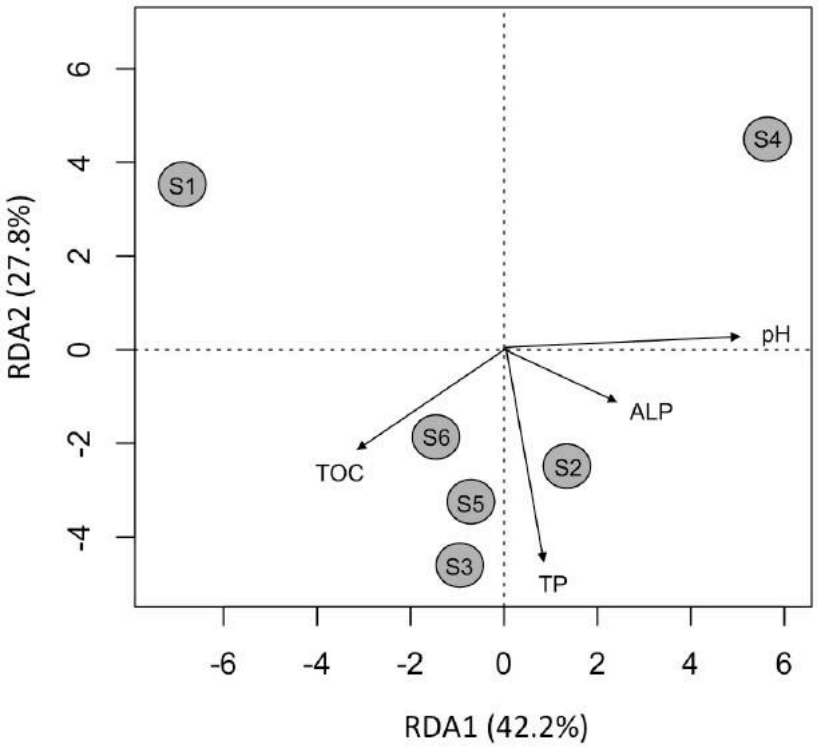


Figure 3.4. Redundancy analysis (RDA) of the *phoX*-harboring community of samples S1 to S6 with the soil properties alkaline phosphatase activity (ALP), total organic C (TOC), total P (TP) and soil pH (pH). Among the tested soil properties, pH and TP were significantly correlated with the *phoX*-harboring community structure.

3.4 Conclusion

The *phoX* gene is widely spread in bacteria and also found in archaea. The meta-analysis showed that the *phoX* gene is found in high frequency in aquatic and terrestrial ecosystems. *phoX*-harboring communities vary greatly between soils, with a pre-dominance of *phoX*-harboring *Actinobacteria* sand *Proteobacteria*. Our results suggest that soil pH and TP influence the *phoX*-harboring community structure. This is the first time that primers targeting *phoX* in soil microorganisms have been designed.

3.5 Acknowledgements

The authors would like to thank Dr. Stefan Zoller for the Perl scripts for taxonomic analysis and the Genetic Diversity Center (Zürich, Switzerland) for technical assistance. We also acknowledge Agroscope (Switzerland) and the New South Wales Department of Primary Industry (NSW, Australia) for access to the sampling sites. This work was supported by the Swiss National Science Foundation (SNF) and by a research grant from the University of Sydney.

3.6 Supplementary material

Figure S3.1. Neighbor-joining tree based on the amino acid sequences of the full length of the *phoX* genes. The tree was rooted using the sequence of the *phoD* alkaline phosphatase gene of *Mesorhizobium loti* MAFF303099 (in black).

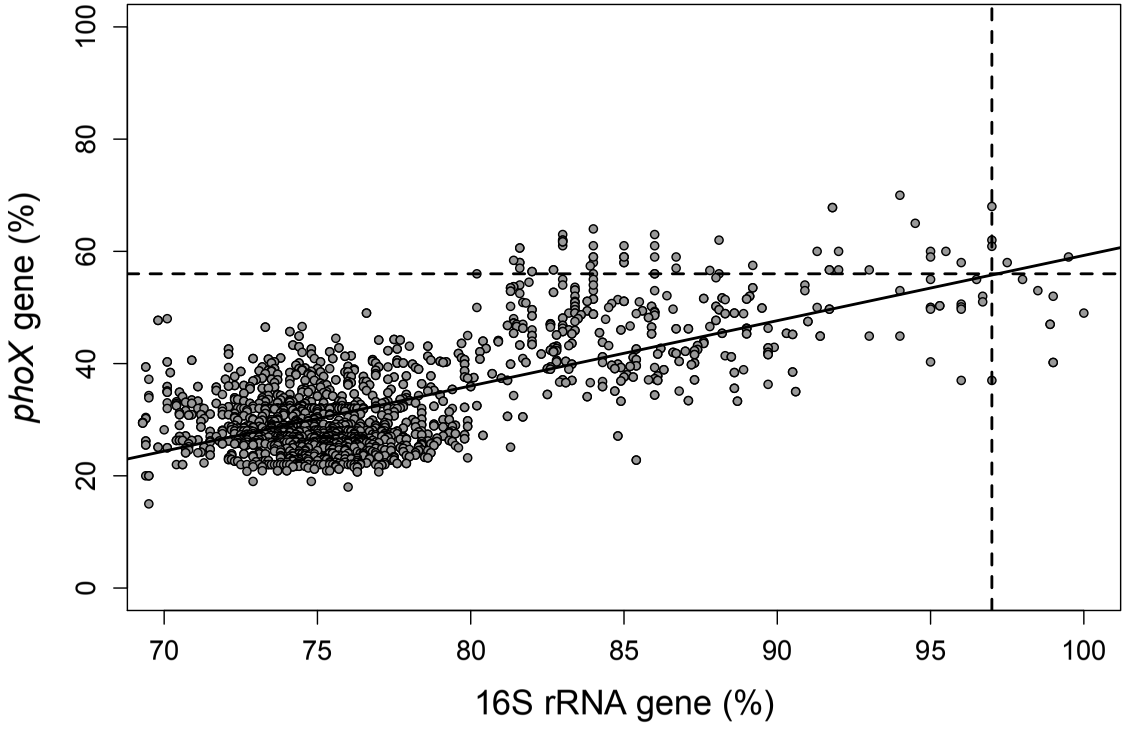


Figure S3.2. Similarity percentages of pairwise alignment of the 16S rRNA gene and *phoX* gene of 20 strains indicated in bold in Table S3.1. Cutoff at 97% similarity of the 16S rRNA gene corresponds to a cutoff at 57% of the *phoX* gene.

Table S3.1. Accession numbers in EBI and NCBI databases of reference sequences used for the primer design targeting the *phoX* gene and taxonomy.

#	Accession number	Phylum	Class	Genus/Species
1	ENA|ACU40818	*Actinobacteria*	*Actinobacteridae*	*Actinosynnema mirum*
2	ENA|ACU40819	*Actinobacteria*	*Actinobacteridae*	*Actinosynnema mirum*
3	ENA|ABM09512	*Actinobacteria*	*Actinobacteridae*	*Arthrobacter aurescens*
4	ENA|ACL40800	*Actinobacteria*	*Actinobacteridae*	*Arthrobacter chlorophenolicus*
5	ENA|ABK04530	*Actinobacteria*	*Actinobacteridae*	*Arthrobacter* sp.
6	ENA|ACQ78939	*Actinobacteria*	*Actinobacteridae*	*Beutenbergia cavernae*
7	ENA|ACU86805	*Actinobacteria*	*Actinobacteridae*	*Brachybacterium faecium*
8	ENA|CAN02934	*Actinobacteria*	*Actinobacteridae*	*Clavibacter michiganensis sub sp.michiganensis*
9	ENA|CAQ03017	*Actinobacteria*	*Actinobacteridae*	*Clavibacter michiganensis sub sp.sepedonicus*
10	ENA|CAI36297	*Actinobacteria*	*Actinobacteridae*	*Corynebacterium jeikeium*
11	ENA|CAQ04307	*Actinobacteria*	*Actinobacteridae*	*Corynebacterium urealyticum*
12	ENA|CAJ64261	*Actinobacteria*	*Actinobacteridae*	*Frankia alni*
13	ENA|CAJ64273	*Actinobacteria*	*Actinobacteridae*	*Frankia alni*
14	**ENA|ABD12759**	***Actinobacteria***	***Actinobacteridae***	***Frankia* sp.**
15	ENA|ABW09955	*Actinobacteria*	*Actinobacteridae*	*Frankia* sp.
16	ENA|ABW10742	*Actinobacteria*	*Actinobacteridae*	*Frankia* sp.
17	ENA|ABW13581	*Actinobacteria*	*Actinobacteridae*	*Frankia* sp.
18	ENA|ABW14921	*Actinobacteria*	*Actinobacteridae*	*Frankia* sp.
19	ENA|ACY23909	*Actinobacteria*	*Actinobacteridae*	*Gordonia bronchialis*
20	ENA|ABS05080	*Actinobacteria*	*Actinobacteridae*	*Kineococcus radiotolerans*
21	ENA|BAG28881	*Actinobacteria*	*Actinobacteridae*	*Kocuria rhizophila*
22	ENA|ACV07128	*Actinobacteria*	*Actinobacteridae*	*Kytococcus sedentarius*
23	ENA|ACV07181	*Actinobacteria*	*Actinobacteridae*	*Kytococcus sedentarius*
24	**ENA|ACS31258**	***Actinobacteria***	***Actinobacteridae***	***Micrococcus luteus***
25	ENA|ABO01519	*Actinobacteria*	*Actinobacteridae*	*Mycobacterium gilvum*
26	ENA|ABG11479	*Actinobacteria*	*Actinobacteridae*	*Mycobacterium* sp.
27	ENA|ABL94653	*Actinobacteria*	*Actinobacteridae*	*Mycobacterium* sp.
28	ENA|ABM16797	*Actinobacteria*	*Actinobacteridae*	*Mycobacterium* sp.
29	ENA|ABP43345	*Actinobacteria*	*Actinobacteridae*	*Mycobacterium vanbaalenii*
30	ENA|ACV80408	*Actinobacteria*	*Actinobacteridae*	*Nakamurella multipartita*
31	ENA|ABL81393	*Actinobacteria*	*Actinobacteridae*	*Nocardioides* sp.
32	ENA|ABG95287	*Actinobacteria*	*Actinobacteridae*	*Rhodococcus erythropolis*
33	ENA|BAH36672	*Actinobacteria*	*Actinobacteridae*	*Rhodococcus jostii*
34	ENA|BAH51524	*Actinobacteria*	*Actinobacteridae*	*Rhodococcus opacus*
35	**ENA|ACU98871**	***Actinobacteria***	***Actinobacteridae***	***Saccharomonospora viridis***
36	ENA|CAM02696	*Actinobacteria*	*Actinobacteridae*	*Saccharopolyspora erythraea*
37	ENA|CAM04233	*Actinobacteria*	*Actinobacteridae*	*Saccharopolyspora erythraea*
38	ENA|CAM06480	*Actinobacteria*	*Actinobacteridae*	*Saccharopolyspora erythraea*
39	ENA|CAM06484	*Actinobacteria*	*Actinobacteridae*	*Saccharopolyspora erythraea*
40	ENA|ABP56001	*Actinobacteria*	*Actinobacteridae*	*Salinispora arenicola*
41	ENA|ABP56479	*Actinobacteria*	*Actinobacteridae*	*Salinispora arenicola*
42	ENA|ABV99739	*Actinobacteria*	*Actinobacteridae*	*Salinispora tropica*
43	ENA|ABW00222	*Actinobacteria*	*Actinobacteridae*	*Salinispora tropica*
44	ENA|BAG18290	*Actinobacteria*	*Actinobacteridae*	*Streptomyces griseus sub sp.griseus*
45	ENA|BAG20625	*Actinobacteria*	*Actinobacteridae*	*Streptomyces griseus sub sp.griseus*
46	ENA|BAG21880	*Actinobacteria*	*Actinobacteridae*	*Streptomyces griseus sub sp.griseus*
47	ENA|AAZ56327	*Actinobacteria*	*Actinobacteridae*	*Thermobifida fusca*

Table S3.1. Accession numbers in EBI and NCBI databases of reference sequences used for the primer design targeting the *phoX* gene and taxonomy. (Continued)

#	Accession number	Phylum	Class	Genus/Species
48	ENA|ABG03285	*Actinobacteria*	*Rubrobacteridae*	*Rubrobacter xylanophilus*
49	**ENA|ACD66365**	***Aquificae***	***Aquificales***	***Sulfurihydrogenibium* sp.**
50	**ENA|ABG60622**	***Bacteroidetes***	***Cytophagia***	***Cytophaga hutchinsonii***
51	ENA|ACL25640	*Chloroflexi*	*Chloroflexales*	*Chloroflexus aggregans*
52	ENA|ABU58082	*Chloroflexi*	*Chloroflexales*	*Roseiflexus castenholzii*
53	ENA|ABQ90106	*Chloroflexi*	*Chloroflexales*	*Roseiflexus* sp.
54	ENA|ABX07063	*Chloroflexi*	*Herpetosiphonales*	*Herpetosiphon aurantiacus*
55	ENA|ABX07063_2	*Chloroflexi*	*Herpetosiphonales*	*Herpetosiphon aurantiacus*
56	ENA|ACM07013	*Chloroflexi*	*Thermomicrobiales*	*Thermomicrobium roseum*
57	ENA|ABA21124	*Cyanobacteria*	*Nostocales*	*Anabaena variabilis*
58	**ENA|ABA21687**	***Cyanobacteria***	***Nostocales***	***Anabaena variabilis***
59	ENA|BAB77409	*Cyanobacteria*	*Nostocales*	*Nostoc punctiforme*
60	ENA|BAB77656	*Cyanobacteria*	*Nostocales*	*Nostoc punctiforme*
61	ENA|ACC80848	*Cyanobacteria*	*Nostocales*	*Nostoc* sp.
62	**ENA|ACC83870**	***Cyanobacteria***	***Nostocales***	***Nostoc* sp.**
63	ENA|BAB76493	*Cyanobacteria*	*Nostocales*	*Nostoc* sp.
64	ENA|ABW25423	*Cyanobacteria*	*Oscillatoriophycideae*	*Acaryochloris marina*
65	ENA|ABW28538	*Cyanobacteria*	*Oscillatoriophycideae*	*Acaryochloris marina*
66	ENA|ACB52926	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
67	ENA|ACB54080	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
68	ENA|ACK66728	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
69	ENA|ACK71581	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
70	ENA|ACK73295	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
71	**ENA|ACL47351**	***Cyanobacteria***	***Oscillatoriophycideae***	***Cyanothece* sp.**
72	ENA|ACV02204	*Cyanobacteria*	*Oscillatoriophycideae*	*Cyanothece* sp.
73	ENA|BAG02841	*Cyanobacteria*	*Oscillatoriophycideae*	*Microcystis aeruginosa*
74	ENA|ABC99852	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
75	ENA|ABD02936	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
76	ENA|ACA98897	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
77	ENA|CAE08314	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
78	ENA|CAK24236	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
79	ENA|CAK29307	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechococcus* sp.
80	ENA|BAD02038	*Cyanobacteria*	*Oscillatoriophycideae*	*Synechocystis* sp.
81	ENA|BAC07717	*Cyanobacteria*	*Oscillatoriophycideae*	*Thermosynechococcus elongatus*
82	ENA|ABG528871	*Cyanobacteria*	*Oscillatoriophycideae*	*Trichodesmium erythraeum*
83	ENA|ABG53451	*Cyanobacteria*	*Oscillatoriophycideae*	*Trichodesmium erythraeum*
84	**ENA|ACO47715**	***Deinococcus-Thermus***	***Deinococci***	***Deinococcus deserti***
85	**ENA|AEN86917**	***Firmicutes***	***Bacilli***	***Bacillus megaterium***
86	ENA|ABQ29501	*Proteobacteria*	*Alphaproteobacteria*	*Acidiphilium cryptum*
87	ENA|AAK890242	*Proteobacteria*	*Alphaproteobacteria*	*Agrobacterium* sp.
88	ENA|ACM27230	*Proteobacteria*	*Alphaproteobacteria*	*Agrobacterium tumefaciens*
89	ENA|ACL94904	*Proteobacteria*	*Alphaproteobacteria*	*Caulobacter crescentus*
90	ENA|ABG62351	*Proteobacteria*	*Alphaproteobacteria*	*Chelativorans* sp.
91	ENA|ABV94113	*Proteobacteria*	*Alphaproteobacteria*	*Dinoroseobacter shibae*
92	ENA|ABI75858	*Proteobacteria*	*Alphaproteobacteria*	*Hyphomonas neptunium*
93	ENA|ABI78805	*Proteobacteria*	*Alphaproteobacteria*	*Hyphomonas neptunium*
94	ENA|ABD54815	*Proteobacteria*	*Alphaproteobacteria*	*Jannaschia* sp.
95	ENA|ABK45595	*Proteobacteria*	*Alphaproteobacteria*	*Magnetococcus marinus*

Table S3.1. Accession numbers in EBI and NCBI databases of reference sequences used for the primer design targeting the *phoX* gene and taxonomy. (Continued)

#	Accession number	Phylum	Class	Genus/Species
96	ENA|ABI65807	*Proteobacteria*	*Alphaproteobacteria*	*Maricaulismaris* sp.
97	ENA|ABY31485	*Proteobacteria*	*Alphaproteobacteria*	*Methylobacterium extorquens*
98	ENA|ACB81441	*Proteobacteria*	*Alphaproteobacteria*	*Methylobacterium extorquens*
99	**ENA|ACK84237**	***Proteobacteria***	***Alphaproteobacteria***	***Methylobacterium extorquens***
100	ENA|ACS410481	*Proteobacteria*	*Alphaproteobacteria*	*Methylobacterium extorquens*
101	ENA|CAX25510	*Proteobacteria*	*Alphaproteobacteria*	*Methylobacterium populi*
102	ENA|ABS17149	*Proteobacteria*	*Alphaproteobacteria*	*Ochrobactrum anthropi*
103	ENA|ABL71545	*Proteobacteria*	*Alphaproteobacteria*	*Paracoccus denitrificans*
104	ENA|ABS64241	*Proteobacteria*	*Alphaproteobacteria*	*Parvibaculumlavamentivorans*
105	ENA|ACG77323	*Proteobacteria*	*Alphaproteobacteria*	*Phenylobacteriumzucineum*
106	ENA|ABN78858	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacter* sp.
107	ENA|ACM03082	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacter sphaeroides*
108	GI|77464988	*Proteobacteria*	*Alphaproteobacteria*	*Rhodobacter sphaeroides*
109	ENA|ABC23801	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillum centenum*
110	ENA|ACI97696	*Proteobacteria*	*Alphaproteobacteria*	*Rhodospirillum rubrum*
111	ENA|ABG30448	*Proteobacteria*	*Alphaproteobacteria*	*Roseobacter denitrificans*
112	ENA|ABI93362	*Proteobacteria*	*Alphaproteobacteria*	*Roseobacter denitrificans*
113	ENA|ABF65287	*Proteobacteria*	*Alphaproteobacteria*	*Ruegeria* sp.
114	ENA|ABR59550	*Proteobacteria*	*Alphaproteobacteria*	*Sinorhizobium fredii*
115	ENA|ACP24635	*Proteobacteria*	*Alphaproteobacteria*	*Sinorhizobium medicae*
116	**ENA|ABQ68833**	***Proteobacteria***	***Alphaproteobacteria***	***Sphingomona swittichii***
117	ENA|EFF75830	*Proteobacteria*	*Betaproteobacteria*	*Achromobacter piechaudii*
118	ENA|ABM330401	*Proteobacteria*	*Betaproteobacteria*	*Acidovorax citrulli*
119	ENA|ABM330401	*Proteobacteria*	*Betaproteobacteria*	*Acidovorax citrulli*
120	ENA|ACM34109	*Proteobacteria*	*Betaproteobacteria*	*Acidovorax ebreus*
121	GI|56475432	*Proteobacteria*	*Betaproteobacteria*	*Aromatoleum aromaticum*
122	ENA|CAL94716	*Proteobacteria*	*Betaproteobacteria*	*Azoarcus* sp.
123	ENA|CAJ48323	*Proteobacteria*	*Betaproteobacteria*	*Bordetella avium*
124	ENA|CAP43811	*Proteobacteria*	*Betaproteobacteria*	*Bordetella petrii*
125	**ENA|ABE35329**	***Proteobacteria***	***Betaproteobacteria***	***Burkholderia multivorans***
126	ENA|ABX15787	*Proteobacteria*	*Betaproteobacteria*	*Burkholderia xenovorans*
127	ENA|ACV35308	*Proteobacteria*	*Betaproteobacteria*	*Candidatus* sp.
128	ENA|ABF10696	*Proteobacteria*	*Betaproteobacteria*	*Cupriavidus metallidurans*
129	ENA|CAP62984	*Proteobacteria*	*Betaproteobacteria*	*Cupriavidus taiwanensis*
130	ENA|ADJ64279	*Proteobacteria*	*Betaproteobacteria*	*Herbaspirillum seropedicae*
131	ENA|ABR90538	*Proteobacteria*	*Betaproteobacteria*	*Janthinobacterium* sp.
132	ENA|ACB33919	*Proteobacteria*	*Betaproteobacteria*	*Leptothrix cholodnii*
133	ENA|ABM94105	*Proteobacteria*	*Betaproteobacteria*	*Methylibium petroleiphilum*
134	ENA|ABM96479	*Proteobacteria*	*Betaproteobacteria*	*Methylibium petroleiphilum*
135	ENA|ABB75861	*Proteobacteria*	*Betaproteobacteria*	*Nitrosospira* sp.
136	ENA|ABE43661	*Proteobacteria*	*Betaproteobacteria*	*Polaromonas* sp.
137	ENA|ABE45115	*Proteobacteria*	*Betaproteobacteria*	*Polaromonas* sp.
138	ENA|CAJ91520	*Proteobacteria*	*Betaproteobacteria*	*Ralstonia eutropha*
139	ENA|ACR01142	*Proteobacteria*	*Betaproteobacteria*	*Thauera* sp.
140	ENA|ACS18868	*Proteobacteria*	*Betaproteobacteria*	*Variovorax paradoxus*
141	ENA|ACS21753	*Proteobacteria*	*Betaproteobacteria*	*Variovorax paradoxus*
142	ENA|ABM57597	*Proteobacteria*	*Betaproteobacteria*	*Verminephrobacter eiseniae*
143	ENA|ACY15956	*Proteobacteria*	*Deltaproteobacteria*	*Haliangium ochraceum*

Table S3.1. Accession numbers in EBI and NCBI databases of reference sequences used for the primer design targeting the *phoX* gene and taxonomy. (Continued)

#	Accession number	Phylum	Class	Genus/Species
144	ENA|CAN91093	*Proteobacteria*	*Deltaproteobacteria*	*Sorangiumc ellulosum*
145	ENA|AAW34736	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacter fetus*
146	ENA|ABK81833	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacter jejuni*
147	ENA|ABS44541	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacter jejuni*
148	ENA|ABV51738	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacter jejuni*
149	ENA|EAQ73176	*Proteobacteria*	*Epsilonproteobacteria*	*Campylobacter jejuni sub sp.jejuni*
150	ENA|ABO10917	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
151	ENA|ACC55785	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
152	ENA|ACJ39967	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
153	ENA|ACJ58853	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
154	ENA|CAM88108	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
155	ENA|CAP02332	*Proteobacteria*	*Gammaproteobacteria*	*Acinetobacter calcoaceticus*
156	ENA|CAL17644	*Proteobacteria*	*Gammaproteobacteria*	*Alcanivorax borkumensis*
157	ENA|ACO79015	*Proteobacteria*	*Gammaproteobacteria*	*Azotobacter vinelandii*
158	ENA|ACE83245	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrio japonicus*
159	ENA|ACE84125	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrio japonicus*
160	ENA|ACE86186	*Proteobacteria*	*Gammaproteobacteria*	*Cellvibrio japonicus*
161	ENA|ABE57951	*Proteobacteria*	*Gammaproteobacteria*	*Chromohalobacte rsalexigens*
162	**ENA|ABV12102**	***Proteobacteria***	***Gammaproteobacteria***	***Citrobacter koseri***
163	ENA|ABU78202	*Proteobacteria*	*Gammaproteobacteria*	*Cronobacter sakazakii*
164	ENA|ACT05442	*Proteobacteria*	*Gammaproteobacteria*	*Dickeyazeae* sp.
165	ENA|ACR70570	*Proteobacteria*	*Gammaproteobacteria*	*Edwardsiella ictaluri*
166	ENA|ABC27041	*Proteobacteria*	*Gammaproteobacteria*	*Hahella chejuensis*
167	ENA|ABC33157	*Proteobacteria*	*Gammaproteobacteria*	*Hahella chejuensis*
168	ENA|ABM19915	*Proteobacteria*	*Gammaproteobacteria*	*Marinobacter aquaeolei*
169	ENA|ABR71219	*Proteobacteria*	*Gammaproteobacteria*	*Marinomonas* sp.
170	ENA|ABR72214	*Proteobacteria*	*Gammaproteobacteria*	*Marinomonas* sp.
171	ENA|ABA58604	*Proteobacteria*	*Gammaproteobacteria*	*Nitrosococcus oceani*
172	ENA|ABG38862	*Proteobacteria*	*Gammaproteobacteria*	*Pseudoalteromonas atlantica*
173	ENA|ABG41122	*Proteobacteria*	*Gammaproteobacteria*	*Pseudoalteromonas atlantica*
174	ENA|AAZ33688	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas entomophila*
175	**ENA|ABA76916**	***Proteobacteria***	***Gammaproteobacteria***	***Pseudomonas fluorescens***
176	ENA|ABJ11858	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas fluorescens*
177	ENA|ABP79623	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas mendocina*
178	ENA|ABP79971	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas mendocina*
179	ENA|ABP85164	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas protegens*
180	ENA|ABP86806	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas putida*
181	**ENA|ABQ77245**	***Proteobacteria***	***Gammaproteobacteria***	***Pseudomonas putida***
182	ENA|ABR83736	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas putida*
183	ENA|ABY96950	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas stutzeri*
184	ENA|ACA74663	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas stutzeri*
185	ENA|CAK16971	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas syringae*
186	ENA|CAW27196	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas syringae*
187	ENA|CAY52915	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas syringae*
188	**ENA|AAO54035**	***Proteobacteria***	***Gammaproteobacteria***	***Pseudomonas aeruginosa***
189	ENA|AAY39717	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas aeruginosa*
190	ENA|AAY94886	*Proteobacteria*	*Gammaproteobacteria*	*Pseudomonas aeruginosa*
191	ENA|ABQ95011	*Proteobacteria*	*Gammaproteobacteria*	*Psychrobacter arcticus*

Table S3.1. Accession numbers in EBI and NCBI databases of reference sequences used for the primer design targeting the *phoX* gene and taxonomy. (Continued)

#	Accession number	Phylum	Class	Genus/Species
192	ENA|EGK08115	*Proteobacteria*	*Gammaproteobacteria*	*Psychrobacter cryohalolentis*
193	ENA|AAZ18170	*Proteobacteria*	*Gammaproteobacteria*	*Psychrobacter* sp.
194	ENA|ABE74115	*Proteobacteria*	*Gammaproteobacteria*	*Psychrobacter* sp.
195	ENA|ABD79649	*Proteobacteria*	*Gammaproteobacteria*	*Saccharophagu sdegradans*
196	ENA|ABV39827	*Proteobacteria*	*Gammaproteobacteria*	*Serratia proteamaculans*
197	ENA|ABK48226	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella amazonensis*
198	ENA|ABL99954	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella baltica*
199	ENA|ABM24820	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella baltica*
200	ENA|ABN61589	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella baltica*
200	ENA|ABN61589	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella baltica*
201	ENAABP75748	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella baltica*
202	ENAABS08402	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella denitrificans*
203	ENAABX49549	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella frigidimarina*
204	ENAACK46585	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella putrefaciens*
205	ENAABE56229	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella* sp.
206	ENAABI39016	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella* sp.
207	ENAABI43022	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella* sp.
208	ENAABI70469	*Proteobacteria*	*Gammaproteobacteria*	*Shewanella* sp.
209	ENAACL71338	*Proteobacteria*	*Gammaproteobacteria*	*Thioalkalivibrio sulfidophilus*
210	ENAABQ19161	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio campbellii*
211	**ENAABU74609**	***Proteobacteria***	***Gammaproteobacteria***	***Vibrio cholerae***
212	ENAACH63840	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio cholerae*
213	ENAACP07016	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio cholerae*
214	ENAACQ62169	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio fischeri*
215	ENAACY53506	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio fischeri*
216	ENAAAW88127	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio* sp.
217	ENABAC97318	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio splendidus*
218	ENACAV26158	*Proteobacteria*	*Gammaproteobacteria*	*Vibrio vulnificus*
219	**ENAAAM36655**	***Proteobacteria***	***Gammaproteobacteria***	***Xanthomonas axonopodis***
220	ENAAAM41064	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas campestris*
221	ENAAAW75611	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas campestris*
222	ENAAAW75715	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas campestris*
223	ENAAAW76390	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas citri*
224	ENAAAY48570	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas euvesicatoria*
225	ENAACD58808	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas oryzae*
226	ENAAGI08395	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas oryzae*
227	ENACAJ24611	*Proteobacteria*	*Gammaproteobacteria*	*Xanthomonas oryzae*
228	ENACAP50896	*Gammaproteobacteria*	*Gammaproteobacteria*	*Xanthomonas oryzae*

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples.

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
Archaea	0	0	0	0	0	0
1. Euryarchaeota	0	0	0	0	0	0
1.1. Methanomicrobia	0	0	0	0	0	0
1.1.1. Halomebacteria	0	0	0	0	0	0
1.1.1.1. Methanothrix	0	0	0	0	0	0
1.1.1.1.1. Methanosaeta	0	0	0	0	0	0
Bacteria	389	2513	568	4694	4504	10489
2. Acidobacteria	0	4	7	0	0	3
2.1. Acidobacteria	0	4	7	0	0	3
2.1.1. Acidobacteria	0	4	7	0	0	3
2.1.1.1. Acidobacteriaceae	0	0	2	0	0	3
2.1.1.1.1. Acidobacterium	0	0	2	0	0	1
2.1.1.1.2. Terriglobus	0	0	0	0	0	2
2.1.1.2. Candidatus	0	4	5	0	0	0
2.1.1.2.1. Chloracidobacterium	0	0	0	0	0	0
2.1.1.2.2. Unclassified Acidobacterium	0	4	5	0	0	0
3. Actinobacteria	9	685	384	1042	879	2312
3.1. Acidimicrobiales	0	0	0	0	0	0
3.1.1. Acidimicrobiaceae	0	0	0	0	0	0
3.1.1.1. Ilumatobacter	0	0	0	0	0	0
3.2. Actinomycetales	9	559	332	1005	820	2163
3.2.1. Actinoplanaceae	0	33	0	123	114	238
3.2.1.1. Actinomyces	0	3	0	0	4	0
3.2.1.2. Actinoplanes	0	8	0	27	35	38
3.2.1.3. Ampullaria	0	4	0	15	0	29
3.2.1.4. Kitasatospora	0	1	0	0	0	0
3.2.1.5. Micromonospora	0	1	0	18	4	29
3.2.1.6. Microstreptospora	0	4	0	32	5	81
3.2.1.7. Salinospora	0	0	0	0	0	1
3.2.1.8. Streptomyces	0	4	0	31	40	28
3.2.1.9. Streptosporangium	0	5	0	0	0	0
3.2.1.10. Streptotrix	0	3	0	0	0	15
3.2.1.11. Thermomonospora	0	0	0	0	26	9
3.2.1.12. Verrucosispora	0	0	0	0	0	8
3.2.2. Actinosynnemataceae	0	111	34	99	24	263
3.2.2.1. Actinobispora	0	0	34	0	0	0
3.2.2.2. Actinosynnema	0	14	0	0	19	23
3.2.2.3. Amycolatopsis	0	3	0	0	0	16
3.2.2.4. Faenia	0	3	0	0	1	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
3.2.2.5. Kutzneria	0	0	0	0	0	0
3.2.2.6. Saccharomonospora	0	0	0	0	0	0
3.2.2.7. Saccharothrix	0	91	0	99	4	224
3.2.2.8. Thermopolyspora	0	0	0	0	0	0
3.2.3. Beutenbergiaceae	0	4	0	1	0	0
3.2.3.2. Beutenbergia	0	4	0	1	0	0
3.2.4. Cellulomonadaceae	0	0	0	6	10	30
3.2.4.1. Aplanobacter	0	0	0	0	0	0
3.2.4.2. Cellulomonas	0	0	0	6	10	30
3.2.5. Corynebacteriaceae	0	0	2	1	0	10
3.2.5.1. Caseobacter	0	0	2	0	0	6
3.2.5.2. Corynebacterium	0	0	0	1	0	4
3.2.5.3. Flavobacterium	0	0	0	0	0	0
3.2.6. Dermabacteraceae	0	0	0	0	1	0
3.2.6.1. Brachybacterium	0	0	0	0	1	0
3.2.7. Dermacoccaceae	0	0	2	0	6	0
3.2.7.1. Dermacoccus	0	0	0	0	0	0
3.2.7.2. Kytococcus	0	0	2	0	6	0
3.2.8. Frankiaceae	0	0	1	0	0	11
3.2.8.1. Frankia	0	0	1	0	0	11
3.2.9. Geodermatophilaceae	0	0	0	0	0	0
3.2.9.1. Geodermatophilus	0	0	0	0	0	0
3.2.10. Gordoniaceae	0	0	0	0	0	0
3.2.10.1. Gordona	0	0	0	0	0	0
3.2.11. Intrasporangiaceae	0	0	0	0	7	0
3.2.11.1. Humihabitans	0	0	0	0	7	0
3.2.12. Kineosporiaceae	0	0	0	3	1	4
3.2.12.1. Kineococcus	0	0	0	3	1	4
3.2.13. Microbacteriaceae	0	0	0	9	8	63
3.2.13.2. Clavibacter	0	0	0	9	8	63
3.2.14. Micrococcaceae	0	36	38	593	40	76
3.2.14.1. Arthrobacter	0	34	38	586	33	58
3.2.14.2. Kocuria	0	1	0	0	0	13
3.2.14.3. Micrococcus	0	1	0	7	7	5
3.2.15. Microsphaeraceae	0	0	0	0	2	79
3.2.15.1. Humicoccus	0	0	0	0	2	79
3.2.16. Mycobacteriaceae	9	367	255	84	436	1250
3.2.16.1. Mycobacterium	9	345	255	5	378	860
3.2.16.2. Nocardia	0	22	0	79	58	390
3.2.17. Nocardiaceae	0	1	0	0	48	24

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
3.2.17.1. Micropolyspora	0	0	0	0	0	0
3.2.17.2. Rhodococcus	0	1	0	0	48	24
3.2.18. Nocardioidaceae	0	4	0	33	2	12
3.2.18.1. Hongia	0	4	0	3	0	0
3.2.18.2. Nocardioides	0	0	0	1	0	3
3.2.18.3. Brachystreptospora	0	0	0	3	2	9
3.2.18.4. Thermonospora	0	0	0	26	0	0
3.2.19. Promicromonosporaceae	0	0	0	0	0	4
3.2.19.1. Isoptericola	0	0	0	0	0	4
3.2.19.2. Xylanimonas	0	0	0	0	0	0
3.2.20. Propionibacteriaceae	0	3	0	53	116	99
3.2.20.1. Microlunatus	0	3	0	53	116	99
3.2.21. Tsukamurellaceae	0	0	0	0	5	0
3.2.21.1. Tsukamurella	0	0	0	0	5	0
3.3. Bifidobacteriales	0	0	0	0	0	0
3.3.1. Bifidibacterium	0	0	0	0	0	0
3.3.1.1. Bifidobacterium	0	0	0	0	0	0
3.4. Rubrobacterales	0	13	50	34	14	80
3.4.1. Rubrobacteraceae	0	13	50	34	14	80
3.4.1.1. Rubrobacter	0	13	50	34	14	80
3.5. Solirubrobacterales	0	113	2	3	45	69
3.5.1. Conexibacteraceae	0	113	2	3	45	69
3.5.1.1. Conexibacter	0	113	2	3	45	69
3.5.1.2. Unclassied Solibacteraceae	0	0	0	0	0	0
4. Aminanaerobia	0	0	0	0	0	0
4.1. Synergistia	0	0	0	0	0	0
4.1.1. Synergistales	0	0	0	0	0	0
4.1.1.1. Anaerobaculum	0	0	0	0	0	0
4.1.1.1.1. Anaerobaculum	0	0	0	0	0	0
5. Armatimonadetes	0	0	0	1	1	14
5.1. Chthonomonadetes	0	0	0	1	0	0
5.1.1. Chthonomonadales	0	0	0	1	0	0
5.1.1.1. Chthonomonadaceae	0	0	0	1	0	0
5.1.1.1.1. Chthonomonas	0	0	0	1	0	0
5.2. Fimbriimonadia	0	0	0	0	1	14
5.2.1. Fimbriionadales	0	0	0	0	1	14
5.2.1.1. Fimbriionadales	0	0	0	0	1	14
5.2.1.1.1. Fimbriimonas	0	0	0	0	1	14
6. Bacteroidetes	0	5	1	0	9	8
6.1. Bacteroidia	0	0	0	0	0	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
6.1.1. Bacteroidales	0	0	0	0	0	0
6.1.1.1. Rhodothermaceae	0	0	0	0	0	0
6.1.1.1.1. Rhodothermus	0	0	0	0	0	0
6.1.1.2. Prolixibacteraceae	0	0	0	0	0	0
6.1.1.2.1. Draconibacterium	0	0	0	0	0	0
6.2. Cytophagia	0	1	0	0	9	2
6.2.1. Cytophagales	0	1	0	0	9	2
6.2.1.1. Cytophagaceae	0	1	0	0	9	2
6.2.1.1.1. Dyadobacter	0	0	0	0	2	1
6.2.1.1.2. Fibrella	0	0	0	0	0	0
6.2.1.1.3. Runella	0	0	0	0	1	1
6.2.1.1.4. Spirosoma	0	0	0	0	6	0
6.2.1.1.5. Taxeobacter	0	1	0	0	0	0
6.2.1.2. Flammeovirgaceae	0	0	0	0	0	0
6.2.1.2.1. Marivirga	0	0	0	0	0	0
6.3. Sphingobacteriia	0	4	1	0	0	6
6.3.1. Sphingobacteriales	0	4	1	0	0	6
6.3.1.1. Chitinophagaceae	0	4	0	0	0	6
6.3.1.1.1. Chitiniphaga	0	0	0	0	0	6
6.3.1.1.2. Niabella	0	4	0	0	0	0
6.3.1.1.3. Niastella	0	0	0	0	0	0
6.3.1.2. Saprospiraceae	0	0	1	0	0	0
6.3.1.2.1. Haliscomenobacter	0	0	1	0	0	0
6.3.1.3. Sphingobacteriaceae	0	0	0	0	0	0
6.3.1.3.1. Pedobacter	0	0	0	0	0	0
7. CandidatusSaccharibacteria	0	0	0	0	0	0
7.1. CandidatusSaccharibacteria	0	0	0	0	0	0
8. Chlorobi	0	0	0	0	0	0
8.1. Chlorobia	0	0	0	0	0	0
8.1.1. Chlorobiales	0	0	0	0	0	0
8.1.1.1. Chlorobiacea	0	0	0	0	0	0
8.1.1.1.1. Chlorobaculum	0	0	0	0	0	0
9. Firmicutes	0	0	0	0	0	0
9.1. Bacilli	0	0	0	0	0	0
9.1.1. Bacillales	0	0	0	0	0	0
9.1.1.1. Bacillaceae	0	0	0	0	0	0
9.1.1.1.1. Bacillus	0	0	0	0	0	0
9.1.1.2. Paenibacillaceae	0	0	0	0	0	0
9.1.1.2.1. Geobacillus	0	0	0	0	0	0
9.1.1.2.2. Thermobacillus	0	0	0	0	0	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
9.2. Clostridia	0	0	14	0	0	0
9.2.1. Clostridiales	0	0	0	0	0	0
9.2.1.1. Clostridiales	0	0	0	0	0	0
9.2.1.1.1. Aerothermobacter	0	0	0	0	0	0
9.2.1.2. Ruminococcaceae	0	0	0	0	0	0
9.2.1.2.1. Ethanoligenens	0	0	0	0	0	0
9.2.2. Halanaerobiales	0	0	0	0	0	0
9.2.2.1. Halanaerobiaceae	0	0	0	0	0	0
9.2.2.1.1. Halothermothrix	0	0	0	0	0	0
9.2.3. Thermoanaerobactergroup	0	0	14	0	0	0
9.2.3.1. Thermoanaerobacteraceae	0	0	14	0	0	0
9.2.3.1.1. Clostridium	0	0	14	0	0	0
7.2. Unclassified	0	0	0	0	0	0
10. Chloroflexi	0	6	50	10	2	19
10.1. Anaerolineae	0	0	0	0	0	0
10.1.1. Anaerolinaeles	0	0	0	0	0	0
10.1.1.1. Anaerolinaceae	0	0	0	0	0	0
10.1.1.1.1. Anaerolinea	0	0	0	0	0	0
10.2. Chloroflexi	0	6	50	10	2	19
10.2.1. Chloroflexaceae	0	6	0	10	2	19
10.2.1.1. Chloroflexaceae	0	2	0	0	0	0
10.2.1.1.1. Chlorocrinis	0	2	0	0	0	0
10.2.1.1.2. Chloroflexus	0	0	0	0	0	0
10.2.1.2. Roseiflexaceae	0	4	0	10	2	19
10.2.1.2.1. Roseiflexus	0	4	0	10	2	19
10.2.2. Herpetosiphonales	0	0	50	0	0	0
10.2.2.1. Herpetosiphonaceae	0	0	50	0	0	0
10.2.2.1.1. Herpetosiphon	0	0	50	0	0	0
10.3. Thermomicrobia	0	0	0	0	0	0
10.3.1. Sphaerobacterales	0	0	0	0	0	0
10.3.1.1. Sphaerobacteraceae	0	0	0	0	0	0
10.3.1.1.1. Sphaerobacter	0	0	0	0	0	0
10.3.2. Thermomicrobiales	0	0	0	11	0	0
10.3.2.1. Thermomicrobiaceae	0	0	0	11	0	0
10.3.2.1.1. Thermomicrobium	0	0	0	11	0	0
11. Cyanobacteria	0	10	0	40	9	32
11.1. CyanobacteriaSubsectionII	0	0	0	0	0	0
11.1.1. Pleurocapsales	0	0	0	0	0	0
11.1.1.1. Chroococcidiopsis	0	0	0	0	0	0
11.2. Gloeobacteria	0	0	0	0	9	32

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
11.2.1. Gloeobacterales	0	0	0	0	9	32
11.2.1.1. Gloeobacter	0	0	0	0	9	32
11.2.2. Synechococcales	0	0	0	0	0	0
11.2.2.1. Synechococcus	0	0	0	0	0	0
11.3. Oscillatoriophycideae	0	10	0	40	0	0
11.3.1. Oscillatoriales	0	10	0	40	0	0
11.3.1.1. Oscillatoriaceae	0	10	0	40	0	0
11.3.1.1.1. Geitlerinema	0	10	0	0	0	0
11.3.1.1.2. Oscillatoria	0	0	0	40	0	0
12. Deinococcus-Thermus	0	0	0	8	1	8
12.1. Hadobacteria	0	0	0	8	1	8
12.1.1. Deinococcales	0	0	0	8	1	8
12.1.1.1. Deinococcaceae	0	0	0	8	1	8
12.1.1.1.1. Deinococcus	0	0	0	8	1	8
12.1.1.2. Trueperaceae	0	0	0	0	0	0
12.1.1.2.1. Truepera	0	0	0	0	0	0
12.1.2. Thermales	0	0	0	0	0	0
12.1.2.1. Thermaceae	0	0	0	0	0	0
12.1.2.1.1. Meiothermus	0	0	0	0	0	0
12.1.2.1.2. Oceanithermus	0	0	0	0	0	0
12.1.2.1.3. Thermus	0	0	0	0	0	0
13. Gemmatimonadetes	0	4	2	0	4	0
13.1. Unclassified Gemmatimonadetes	0	4	2	0	4	0
14. Nitrospirae	0	0	0	0	0	0
14.1. Nitrospira	0	0	0	0	0	0
14.1.1. Nitrospirales	0	0	0	0	0	0
14.1.1.1. Nitrospiraceae	0	0	0	0	0	0
14.1.1.1.1. Nitrospira	0	0	0	0	0	0
15. Planctomycetes	0	3	25	41	76	404
15.1. Phycisphaerae	0	0	0	0	0	0
15.1.1. Phycisphaerales	0	0	0	0	0	0
15.1.1.1. Phycisphaeraceae	0	0	0	0	0	0
15.1.1.1.1. Phycisphaera	0	0	0	0	0	0
16.2. Planctomycetacia	0	3	25	41	76	404
16.2.1. Planctomycetales	0	3	25	41	76	404
16.2.1.1. Planctomycetaceae	0	3	25	41	76	404
16.2.1.1.1. Isosphaera	0	3	25	41	73	386
16.2.1.1.2. Pirellula	0	0	0	0	0	6
16.2.1.1.3. Planctomyces	0	0	0	0	3	12
16.2.1.1.4. Singulisphaera	0	0	0	0	0	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
16.3. Unclassified Planctomycetales	0	0	0	0	0	0
16. Proteobacteria	369	1786	93	3552	3474	7689
16.1. Alphaproteobacteria	162	1437	26	3181	3020	7285
16.1.1. Caulobacterales	7	3	0	0	2	0
16.1.1.1. Caulobacteraceae	7	3	0	0	2	0
16.1.1.1.1. Asticcacaulis	0	0	0	0	0	0
16.1.1.1.2. Brevundimonas	0	2	0	0	0	0
16.1.1.1.3. Caulobacter	7	1	0	0	2	0
16.1.1.1.4. Phenylobacterium	0	0	0	0	0	0
16.1.2. Rhizobiales	144	606	22	1155	2214	3856
16.1.2.1. Hyphomicrobiaceae	0	210	0	533	311	578
16.1.2.1.1. Devosia	0	2	0	7	7	26
16.1.2.1.2. Hyphomicrobium	0	208	0	526	304	552
16.1.2.1.3. Rhodomicrobium	0	0	0	0	0	0
16.1.2.2. Beijerinckiaceae	0	2	0	0	0	0
16.1.2.2.1. Methylocella	0	2	0	0	0	0
16.1.2.3. Bradyrhizobiaceae	0	9	0	0	0	2
16.1.2.3.1. Agromonas	0	4	0	0	0	0
16.1.2.3.2. Bradyrhizobium	0	3	0	0	0	2
16.1.2.3.3. Nitrobacter	0	0	0	0	0	0
16.1.2.3.4. Oligotropha	0	0	0	0	0	0
16.1.2.3.5. Rhizobacterium	0	0	0	0	0	0
16.1.2.3.6. Rhodopseudomonas	0	2	0	0	0	0
16.1.2.4. Brucellaceae	0	0	0	2	0	15
16.1.2.4.1. Ochrobactrum	0	0	0	2	0	15
16.1.2.5. Methylobacteriaceae	0	125	8	331	165	392
16.1.2.5.1. Methylobacterium	0	125	8	331	165	392
16.1.2.6. Methylocystaceae	0	0	0	0	0	0
16.1.2.6.1. Methylocystis	0	0	0	0	0	0
16.1.2.7. Phyllobacteriaceae	14	68	10	67	625	1832
16.1.2.7.1. Chelativorans	0	5	0	33	12	36
16.1.2.7.2. Mesorhizobium	14	63	10	34	613	1796
16.1.2.8. Rhizobiaceae	130	172	4	199	1014	810
16.1.2.8.1. Agrobacterium	0	37	0	12	19	235
16.1.2.8.2. Ensifer	0	4	0	0	0	1
16.1.2.8.3. Rhizobium	42	95	4	57	783	372
16.1.2.8.4. Sinorhizobium	88	36	0	130	212	202
16.1.2.9. Xanthobacteraceae	0	20	0	23	99	227
16.1.2.9.1. Unclassified Xanthobacteraceae	0	20	0	23	99	227
16.1.3. Rhodobacterales	0	446	1	707	292	1918

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
16.1.3.1. Rhodobacteraceae	0	446	1	707	292	1918
16.1.3.1.1. Dinoroseobacter	0	74	0	226	160	768
16.1.3.1.2. Erythrobacter	0	2	0	0	2	19
16.1.3.1.3. Leisingera	0	133	0	190	49	142
16.1.3.1.4. Paracoccus	0	224	1	262	78	962
16.1.3.1.5. Rhodobacter	0	12	0	28	3	13
16.1.3.1.6. Roseibacterium	0	0	0	0	0	0
16.1.3.1.7. Roseobacter	0	0	0	0	0	0
16.1.3.1.8. Ruegeria	0	0	0	0	0	14
16.1.3.1.9. Unclassified Rhodobiaceae	0	1	0	1	0	0
16.1.4. Rhodospirillales	11	382	1	1318	512	1511
16.1.4.1. Acetobacteraceae	0	0	0	0	0	0
16.1.4.1.1. Acidiphilium	0	0	0	0	0	0
16.1.4.2. Rhodospirillaceae	11	382	1	1318	512	1511
16.1.4.2.1. Azospirillum	0	374	1	1265	483	1297
16.1.4.2.2. Magnetospirillum	0	3	0	0	2	5
16.1.4.2.3. Rhodocista	0	2	0	6	1	117
16.1.4.2.4. Rhodospirillum	11	3	0	47	26	89
16.1.4.2.5. Thalassospira	0	0	0	0	0	3
16.1.4.2.6. Tistrella	0	0	0	0	0	0
16.1.5. Sphingomonadales	0	0	2	1	0	0
16.1.5.1. Sphingomonadaceae	0	0	2	1	0	0
16.1.5.1.1. Sphingobium	0	0	0	0	0	0
16.1.5.1.2. Sphingomonas	0	0	2	1	0	0
16.1.5.1.3. Sphingopyxis	0	0	0	0	0	0
16.2. Betaproteobacteria	9	10	39	1	128	74
16.2.1. Burkholderiales	0	1	39	1	128	69
16.2.1.1. Alcaligenaceae	0	0	0	0	0	0
16.2.1.1.1. Castellaniella	0	0	0	0	0	0
16.2.1.2. Burkholderiaceae	0	1	2	0	4	3
16.2.1.2.1. Acinetobacter	0	0	0	0	0	0
16.2.1.2.2. ATCC	0	0	0	0	0	0
16.2.1.2.3. Burkholderia	0	1	1	0	0	3
16.2.1.2.4. Cupriavidus	0	0	1	0	0	0
16.2.1.2.5. Ralstonia	0	0	0	0	4	0
16.2.1.3. Comamonadaceae	0	0	37	1	124	66
16.2.1.3.1. Acidivorax	0	0	0	0	2	0
16.2.1.3.2. Acidovorax	0	0	0	0	0	0
16.2.1.3.3. Alicycliphilus	0	0	37	0	0	0
16.2.1.3.4. Comamonadaceae	0	0	0	0	0	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
16.2.1.3.5. Delftia	0	0	0	0	0	0
16.2.1.3.6. Leptothrix	0	0	0	1	93	66
16.2.1.3.7. Polaromonas	0	0	0	0	0	0
16.2.1.3.8. Variovorax	0	0	0	0	29	0
16.2.1.4. Oxalobacteraceae	0	0	0	0	0	0
16.2.1.4.1. Collimonas	0	0	0	0	0	0
16.2.1.4.2. Janthinobacterium	0	0	0	0	0	0
16.2.1.5. Rubrivivax	0	0	0	0	0	0
16.2.1.5.1. Rubrivivax	0	0	0	0	0	0
16.2.1.6. Burkholderiales	0	0	0	0	0	0
16.2.1.6.1. Burkholderiales	0	0	0	0	0	0
16.2.2. Gallionellales	9	0	0	0	0	0
16.2.2.1. Gallionella	9	0	0	0	0	0
16.2.2.1.1. Siderooxidans	9	0	0	0	0	0
16.2.3. Neisseriales	0	0	0	0	0	0
16.2.3.1. Chromobacteriaceae	0	0	0	0	0	0
16.2.3.1.1. Pseudogulbenkiania	0	0	0	0	0	0
16.2.4. Nitrosomonadales	0	0	0	0	0	0
16.2.4.1. Nitrosomonadaceae	0	0	0	0	0	0
16.2.4.1.1. Nitrosomonas	0	0	0	0	0	0
16.2.4.1.2. Nitrosospira	0	0	0	0	0	0
16.2.5. Rhodocyclales	0	9	0	0	0	5
16.2.5.1. Rhodocyclaceae	0	9	0	0	0	5
16.2.5.1.1. Aromatoleum	0	0	0	0	0	0
16.2.5.1.2. Azoarcus	0	0	0	0	0	5
16.2.5.1.3. Thauera	0	9	0	0	0	0
16.2.6. Sulfuricellales	0	0	0	0	0	0
16.2.6.1. Sulfuricellaceae	0	0	0	0	0	0
16.2.6.1.1. Sulfuricella	0	0	0	0	0	0
16.2.7. Unclassified Betaproteobacteria	0	0	0	0	0	0
16.2.7.1. Candidatus	0	0	0	0	0	0
16.2.7.1.1. Candidatus-Accumulibacter	0	0	0	0	0	0
16.3. Deltaproteobacteria	197	6	0	0	3	2
16.3.1. Desulfuromonadales	197	2	0	0	2	0
16.3.1.1. Geobacteraceae	196	0	0	0	2	0
16.3.1.1.1. Geobacter	196	0	0	0	2	0
16.3.1.2. Pelobacteraceae	0	0	0	0	0	0
16.3.1.2.1. Pelobacter	0	0	0	0	0	0
16.3.1.3. Anaeromyxobacteraceae	0	0	0	0	0	0
16.3.1.3.1. Anaeromyxobacter	0	0	0	0	0	0

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
16.3.1.4. Archangiaceae	1	2	0	0	0	0
16.3.1.4.1. Stigmatella	1	2	0	0	0	0
16.3.1.5. Haliangiaceae	0	0	0	0	0	0
16.3.1.5.1. Haliangium	0	0	0	0	0	0
16.3.2. Myxococcales	0	4	0	0	1	2
16.3.2.1. Myxococcaceae	0	0	0	0	0	2
16.3.2.1.1. Chondrococcus	0	0	0	0	0	0
16.3.2.1.2. Myxococcus	0	0	0	0	0	2
16.3.2.2. Sorangiaceae	0	4	0	0	0	0
16.3.2.2.1. Sorangium	0	4	0	0	0	0
16.3.2.3. Syntrophobacteraceae	0	0	0	0	1	0
16.3.2.3.1. Syntrophobacter	0	0	0	0	1	0
16.4. Gammaproteobacteria	1	333	28	370	323	328
16.4.1. Aeromonadales	0	0	0	0	0	8
16.4.1.1. Aeromonadaceae	0	0	0	0	0	8
16.4.1.1.1. Aeromonas	0	0	0	0	0	8
16.4.1.2. Alteromonadaceae	0	0	0	0	0	0
16.4.1.2.1. Marinobacter	0	0	0	0	0	0
16.4.1.2.2. Microbulbifer	0	0	0	0	0	0
16.4.2. Chromatiales	0	0	0	0	1	2
16.4.2.1. Chromatiaceae	0	0	0	0	1	0
16.4.2.1.1. Allochromatium	0	0	0	0	0	0
16.4.2.1.2. Nitrosococcus	0	0	0	0	1	0
16.4.2.1.3. Thioflavicoccus	0	0	0	0	0	0
16.4.2.2. Ectothiorhodospira	0	0	0	0	0	2
16.4.2.2.1. Alkalilimnicola	0	0	0	0	0	0
16.4.2.2.2. Thioalkalivibrio	0	0	0	0	0	2
16.4.3. Enterobacteriales	0	48	22	4	17	5
16.4.3.1. Enterobacteraceae	0	48	22	4	17	5
16.4.3.1.1. Aerobacter	0	0	0	0	0	1
16.4.3.1.2. Erwinia	0	0	22	0	0	0
16.4.3.1.3. Pantoea	0	48	0	0	0	4
16.4.3.1.4. Serratia	0	0	0	4	17	0
16.4.3.1.5. Unclassified Enterobacteraceae	0	0	0	0	0	0
16.4.4. Oceanospirillales	0	22	5	0	0	0
16.4.4.1. Halomonadaceae	0	0	0	0	0	0
16.4.4.1.1. Chromohalobacter	0	0	0	0	0	0
16.4.4.2. Hahellaceae	0	22	5	0	0	0
16.4.4.2.1. Hahella	0	22	5	0	0	0
16.4.5. Pseudomonadales	0	256	0	366	305	313

Table S3.2. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	S1	S2	S3	S4	S5	S6
Total	389	2513	568	4694	4504	10489
16.4.5.1. Pseudomonadaceae	0	256	0	366	305	313
16.4.5.1.1. Azotobacter	0	0	0	0	0	0
16.4.5.1.2. Chlorobacterium	0	26	0	0	0	0
16.4.5.1.3. Pseudomonas	0	230	0	366	305	313
16.4.6. Xanthomonadales	1	7	1	0	0	0
16.4.6.1. Lysobacteraceae	1	7	1	0	0	0
16.4.6.1.1. Phytomonas	1	7	0	0	0	0
16.4.6.1.2. Pseudoxanthomonas	0	0	1	0	0	0
16.4.6.1.3. Rhodanobacter	0	0	0	0	0	0
16.4.6.1.4. Stenotrophomonas	0	0	0	0	0	0
17. Spirochaetae	0	0	0	0	0	0
17.1. Spirochaetales	0	0	0	0	0	0
17.1.1. Leptospiraceae	0	0	0	0	0	0
17.1.1.1. Leptospira	0	0	0	0	0	0
18. Verrucomicrobia	4	4	1	0	22	0
18.1. Opitutales	4	4	1	0	20	0
18.1.1. Opitutaceae	4	4	1	0	20	0
18.1.1.1. Opitutus	4	4	0	0	9	0
18.1.1.2. Unclassified Opitutus	0	0	1	0	1	0
18.3. Verrucomicrobiales	0	0	0	0	2	0
18.3.1 Verrucomicrobia	0	0	0	0	2	0
18.3.1.1 Akkermansia	0	0	0	0	2	0
19. Unclassified	7	6	5	0	27	0

Table S3.3. Coverage test of the published and newly-designed *phoX*-targeting primers tested against the 228 sequences used as references in this study.

Primer set	Nb of sequences targeted by the forward primer	Nb of sequences targeted by the reverse primer	Nb of sequences targeted by the primer set in in-silico PCR[1]	Potential fragment size (bp) size (bp)	Primers published in
*phoX*1	13	3	7	682-756	Sebastian and Ammerman (2009)
*phoX*2	26	20	41	647-674	Sebastian and Ammerman (2009)
*phoX*3	19	21	35	497-535	Sebastian and Ammerman (2009)
phoX-F455/R1076	61	48	103	545-615	This study

[1]In-Silico PCR included a potential mismatch per primers.

4

SOIL PHOD AND PHOX ALKALINE PHOSPHATASE GENE COMPOSITION RESPOND TO MULTIPLE ENVIRONMENTAL FACTORS

Combined with Chapter 3 and submitted as:

Ragot SA, MA Kertesz, É Mészáros, E Frossard and EK Bünemann. Soil *phoD* and *phoX* alkaline phosphatase gene compositions respond to multiple environmental factors. *FEMS Microbiology Ecology*.

Abstract

Alkaline phosphatases such as PhoD and PhoX are key players in organic phosphorus cycling in soil. We identified the key organisms harboring the *phoD* and *phoX* gene in soil and explored the relationships between environmental factors and the *phoD*- and *phoX*-harboring community structures across three land-uses located in arid to temperate climates on two continents using 454-sequencing. The *phoD* and *phoX* gene were investigated using the primers designed in Chapter 2 and 3, respectively.

The *phoD* gene was found in 1 archaeal, 13 bacterial and 2 fungal phyla, and the *phoX* gene in 1 archaeal and 16 bacterial phyla. Dominant *phoD*-harboring phyla were *Actinobacteria*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes* and *Proteobacteria*, while abundant *phoX*-harboring phyla were *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Planctomycetes*, *Proteobacteria* and *Verrucomicrobia*. Climate, soil group, land-use and soil nutrient concentrations were the common environmental drivers of the *phoD*- and *phoX*-harboring community structures. In addition, the *phoX*-harboring community structure was affected by soil pH.

Despite differences in environmental factors, dominant phyla in the *phoD*-harboring community remained similar in all samples, while the *phoX*-harboring community composition differed substantially between the samples. This study shows that *phoD* and *phoX* are mostly governed by the same environmental drivers but occur in part in different key phyla.

4.1 Introduction

Soil microorganisms play an important role in cycling of organic phosphorus (P) by producing phosphatases, i.e. enzymes that hydrolyze organic P to orthophosphate, the only bioavailable P form for plants and microorganisms in soil (Frossard et al. 1995; Condron et al. 2005). Phosphatases are produced by many terrestrial organisms when facing P scarcity, including archaea, bacteria, fungi and plants. In bacteria, phosphatase genes are part of the Pho regulon, a global regulatory network involved in P management, which includes functional genes coding for phosphatases such as acid and alkaline phosphatases, high-affinity phosphate transporters, and other systems for mobilizing P (Santos-Beneit 2015). Alkaline phosphatases are common in the environment and particularly widespread across the bacterial kingdom but have also been found in archaea and fungi (Ragot et al. 2015). PhoD and PhoX alkaline phosphatases are monomeric enzymes that catalyze the hydrolysis of both phosphomonoesters and phosphodiesters (Zaheer et al. 2009; Kageyama et al. 2011), which can represent up to 90% of the organic P fraction in soils (Condron et al. 2005). Since a lack of available orthophosphate limits plant productivity in many terrestrial ecosystems, the production of microbial phosphatases plays an important role in plant nutrition (Richardson and Simpson 2011). However, our knowledge of how the environment affects these microbial functional genes remains very limited.

In soil microorganisms, the *phoD* gene is mainly found in the bacterial phyla *Actinobacteria*, *Cyanobacteria*, *Deinococcus-Thermus*, *Planctomycetes* and *Proteobacteria* (Tan et al. 2013; Ragot et al. 2015). Water-soluble phosphate addition has been reported to increase (Tan et al. 2013), reduce (Chhabra et al. 2013) or not affect (Jorquera et al. 2014) the diversity of the *phoD* gene. In arable and grassland soils, pH was found to be a key driver of the total *phoD*-harboring community structure (Wang et al. 2012b; Cui et al. 2015; Ragot et al. 2015), although this is not clear-cut, since in some of these studies, organic carbon (C) or available P also increased together with pH (Cui et al. 2015; Wang

et al. 2012b). Vegetation has also been reported to change the *phoD*-harboring community structure in monoculture and in mixed culture of wheat and legumes (Wang et al. 2012a). In contrast to *phoD*, our knowledge of *phoX* in soil is poor. Metagenomic studies of the Global Ocean Sampling (GOS) gave the first insights into the prevalence and composition of the *phoX* gene in the marine environment (Luo et al. 2009; Sebastián and Ammerman 2009). Sebastián and Ammerman (2009) designed the first primers targeting the *phoX* gene based on sequences of marine bacteria such as *Vibrio* sp., *Shewanella* sp. and *Roseobacter* sp. They showed that *phoX* is common both in the Sargasso Sea and in the Chesapeake Bay. In marine bacteria, the *phoX* gene is most frequently found in *Proteobacteria* and *Cyanobacteria*, but has also been found in *Actinobacteria*, *Bacteroidetes*, *Chloroflexi* and *Lentisphaerae* (Luo et al. 2009; Sebastián and Ammerman 2009). The *phoX* gene was also found in a few bacteria isolated from soil such as *Sinorhizobium meliloti* (Zaheer et al. 2009). On the whole, however, comprehensive knowledge on the effect of environmental factors on *phoD* and *phoX* genes in soil is still missing, especially on a large biogeographical scale.

In this study, we identify the key organisms harboring the *phoD* and *phoX* gene in soil and explore how the *phoD*- and *phoX*-harboring community structure and composition are affected by selected environmental factors. The *phoD* and *phoX* genes were investigated using primers designed in Chapter 2 and 3, respectively. We sampled across three land-uses (arable, grassland and forest) on two continents and assessed the relative importance of climate, soil physical, chemical and biological properties as drivers of the *phoD*- and *phoX*-harboring community structure and composition in a gene-targeted metagenomic approach using 454-sequencing. We hypothesized that the *phoD*- and *phoX*-harboring community structure and composition are affected (i) by climate and land-use which are strong drivers of the total bacterial and fungal communities (Drenovsky et al. 2004); (ii) by P availability assuming that low P availability promotes specific *phoD*- and *phoX*-harboring microorganisms, resulting in an uneven community composed by few dominant

phyla and (iii) by soil pH as both PhoD and PhoX reach their optimum hydrolyzing activity at alkaline pH.

4.2 Material and Methods

4.2.1 Site description and soil sampling

Soil was collected from five locations in Australia in spring 2013 and from five locations in Switzerland in summer 2014 (Table 4.1), covering a broad range of soil groups, vegetation and climatic conditions. Australian soils were sampled in three climate zones, from arid climate (Bsh) to warm temperate climate (Cfa and Cfb), while the Swiss sites were all classified as warm temperate climate (Cfb) (Kottek et al. 2006).

At each location, soil from three land-uses, namely arable, forest and grassland, was collected, except at two Australian sites, where arable fields were missing. We assumed that all arable soils were fertilized, while all grassland soils were presumably not fertilized unless mentioned otherwise. The litter was removed in forest soils before sampling. The dominant plant species were identified at each site (Table 4.1). For each sampling, five soil cores from the top 5 cm were randomly collected and homogenized by sieving (4 mm). Subsamples were stored at -20°C for molecular analysis and at 4°C for microbial biomass analysis. The remaining composite soil was air-dried and used to determine basic soil properties.

4.2.2 Basic soil properties

4.2.2.1 Soil texture, carbonate content, total C, N and P, and pH

Texture and carbonate content ($CaCO_3$-C) were determined by a commercial soil analysis laboratory (Sol-Conseil, Gland, Switzerland). Total C (TC) and N (TN) in soil were

Table 4.1. General sampling site characteristics.

#	Sample	Location	Land-use	Geographical coordinates	Climate[1]	Soil group[2]	Vegetation
1	AUS1-A	Glenroy, Australia	Arable	34°35'6"S, 148°38'5"E	Cfb	Planosol	*Triticum* sp.
2	AUS1-F		Forest	34°40'4"S, 148°38'4" E	Cfb	Planosol	*Vulpia bromoides*, *Briza maxima* *Eucaliptus blackelyi*, *Eucalyptus macrorhyncha*,
3	AUS1-Gf		Fertilized grassland	34°35'6"S, 148°38'5"E	Cfb	Planosol	*Microlaena stipoides*, *Austrodanthonia* sp., *Elymus scaber*, *Themada triandra*
4	AUS1-G		Grassland	34°35'6"S, 148°38'5"E	Cfb	Planosol	*Microlaena stipoides*, *Austrodanthonia* sp., *Elymus scaber, Themada triandra*
5	AUS2-F	Kia-Ora, Australia	Forest	34°48'18"S, 148°35'0"E	Cfb	Planosol	*Eucalyptus albens*
6	AUS2-Gf		Fertilized grassland	34°48'18"S, 148°35'0"E	Cfb	Planosol	*Microlaena stipoides, Austrodanthonia* sp., *Elymus scaber*
7	AUS2-G		Grassland	34°48'18"S, 148°35'0"E	Cfb	Planosol	*Microlaena stipoides, Austrodanthonia* sp., *Elymus scaber*
8	AUS3-A	Narrabi, Australia	Arable	30°16'22"S, 149°48'20"E	Cfa	Planosol	*Triticum* sp.
9	AUS3-F		Forest	30°15'44"S, 149°50'12"E	Cfa	Vertisol	*Enneapogon nigricans, Casuarina* sp., *Eucalyptus melliodora, Callitris glaucephylla*
10	AUS3-G		Grassland	30°15'14"S, 149°51'53"E	Cfa	Planosol	*Chrysocephalum* sp., *Themada* sp., *Festuca arundinacea*
11	AUS4-A	Nyngan, Australia	Arable	31°25'53"S, 147°3'56"E	BSh	Cambisol	*Triticum* sp.
12	AUS4-F		Forest	31°31'21"S, 147°8'48"E	BSh	Cambisol	*Acacia dealbata, Acacia longifolia*
13	AUS4-G		Grassland	31°25'52"S, 147°4'9"E	BSh	Cambisol	Mixed grasses and dicot plants
14	AUS5-F	Mutawintji, Australia	Forest	31°16'5"S, 142°17'22"E	BSh	Leptosol	*Acacia dealbata, Acacia aneura*, *Acacia longifolia*
15	AUS5-G		Grassland	31°16'19"S, 142°17'44"E	BSh	Leptosol	Mixed grasses and dicot plants
16	CH1-A	Basel, Switzerland	Arable	47°30'10"N, 7°32'21"E	Cfb	Luvisol	*Triticum* sp.[3]
17	CH1-F		Forest	47°30'29"N, 7°32'16"E	Cfb	Luvisol	*Fagus silvatica, Galium rotundifolium*, *Maianthemum bifolium*
18	CH1-G		Grassland	47°30'9"N, 7°32'25"E	Cfb	Luvisol	Winter ryegrass, Dactylis glomerata, *Festuca rubra, Trifolium pratense*
19	CH2-A	Eschikon, Switzerland	Arable	47°26'32"N, 8°40'50"E	Cfb	Cambisol	*Brassica napus*
20	CH2-F		Forest	47°26'32"N, 8°40'51"E	Cfb	Cambisol	*Fagus silvatica, Hedera helix,*
21	CH2-G		Grassland	47°26'48"N, 8°40'40"E	Cfb	Cambisol	*Lolium perenne, Dactylis glomerata*, *Phleum pratense*
22	CH3-A	Reckenholz, Switzerland	Arable	47°25'50"N, 8°31'17"E	Cfb	Cambisol	*Triticum* sp.
23	CH3-F		Forest	47°25'44"N, 8°31'32"E	Cfb	Cambisol	*Picea abies, Fagus sylvatica,* *Acer pseudoplatanus*
24	CH3-G		Grassland	47°25'51"N, 8°31'17"E	Cfb	Cambisol	*Agrostis capillaris, Plantago lanceolata,* *Lolium perenne*
25	CH4-A	Watt Switzerland	Arable	47°25'45"N, 8°29'35"E	Cfb	Cambisol	*Triticum* sp.
26	CH4-F		Forest	47°25'47"N, 8°29'34"E	Cfb	Cambisol	*Fagus silvatica, Acer pseudoplatanus*
27	CH4-G		Grassland	47°25'45"N, 8°29'31"E	Cfb	Cambisol	*Arrhenatherum elatius, Ranunculus acris,* *Anthoxanthum odoratum, Holcus lanatus*
28	CH5-A	Reckenholz, Switzerland	Arable	47°25'35"N, 8°31'9"E	Cfb	Luvisol	*Triticum* sp.
29	CH5-F		Forest	47°25'36"N, 8°31'11"E	Cfb	Luvisol	*Fagus silvatica, Quercus rubra*
30	CH5-G		Grassland	47°25'36"N, 8°31'9"E	Cfb	Luvisol	*Anthoxanthum odoratum, Dactylis glomerata,* *Ranunculus acris*

[1] Köppen climate classification (Kottek et al. 2006) (BSh: Arid climate, hot steppe, Cfa: Warm temperate climate, fully humid with hot summer, Cfb: Warm temperate climate, fully humid with warm summer).
[2] World Reference Base for Soil Resources (IUSS 2014).
[3] Mäder et al. (2002).

measured on dry and ground soil using a CNS analyzer (Thermo-Finnigan, CA). Total organic C (TOC) was calculated by subtracting $CaCO_3$-C from TC. Total P (TP) in soil was determined by wet digestion with H_2O_2/H_2SO_4 (Anderson and Ingram 1993) and measured with malachite green at 610 nm (Ohno and Zibilske 1991). Soil pH was measured in a soil suspension in 0.01M $CaCl_2$ in a 1 to 2.5 (w/v) ratio (shaken for 1 hour) using a Benchtop pH 720A (Orion Research Inc., Jacksonville, FL).

4.2.2.2 NaOH-EDTA extractable organic P and resin-extractable P

Organic P (P_{org}) was determined using alkaline extraction after Bowman and Moir (1993). Briefly, 2.5 g of dry and ground soil were extracted with 0.25 M NaOH - 0.05 M EDTA in a 1 to 10 (w/v) ratio for 16 h on a horizontal shaker. Soil extracts were then centrifuged (10 min at ca. 2100 x g) and filtered (Whatman No. 1, Sigma-Aldrich Chemie GmBH, Buchs, Switzerland). The concentration of inorganic P in diluted extracts was measured colorimetrically using malachite green. The concentration of total P in the extracts was measured colorimetrically after digestion with a digestion mix (prepared with 6 g ammonium persulphate dissolved in 100 mL of 0.9 M H_2SO_4) in an autoclave (20 min, 121°C) and neutralization. P_{org} was calculated as the difference between total and inorganic NaOH-EDTA extractable P. Resin-extractable P (P_{res}) was determined after Kouno et al. (1995) as modified by Bünemann et al. (2007) using anion-exchange resin membranes (BDH laboratory supplies #55164 2S, Poole, UK) charged with CO_3^{2-}. Briefly, 2 g dry soil was extracted in 30 mL ddH_2O together with a resin membrane (2 cm x 6 cm) for 16 h at 160 min^{-1} on a horizontal shaker. Resins were eluted in 0.1 M NaCl/0.1 M HCl for 2 h and extracted P was measured colorimetrically using malachite green.

4.2.2.3 Microbial C and N

Microbial C and N analysis was carried out on moist soil using chloroform fumigation-extraction Vance et al. (1987). TOC and TN in the extracts were measured using

a Formacs^{SERIES} TOC/TN analyzer (Skalar, Breda, Netherlands). Concentrations in non-fumigated samples were subtracted from those in fumigated samples to calculate chloroform-labile C (C_{mic}) and N (N_{mic}), without the use of a conversion factor, since it is soil-specific and has not been determined for the soils in our study (Oberson and Joner 2005).

4.2.2.4 Potential acid and alkaline phosphatase activity

Potential acid and alkaline phosphatase activities were determined following Marx et al. (2001) as modified by Poll et al. (2006). Briefly, 1 g of fresh soil was dispersed in 100 mL of autoclaved H_2O using an ultrasonic probe (Labsonic U, Braun Melsungen, Germany) for 50 J s^{-1} for 120 s. The assay was carried out on 50 µL aliquots of the soil suspension in a microplate with 6 analytical replicates, using 4-methylumbelliferylphosphate as substrate, and 0.1 M MES buffer (pH 6.1) (Marx et al. 2001) or Modified Universal Buffer (pH 11) (Alef and Nannipieri 1995) as buffer for acid and alkaline phosphatase activity, respectively. Microplates were incubated for 3 h at 30°C on a horizontal shaker. The increase in fluorescence over time was measured at 360/460 nm and converted into nmol substrate g $soil^{-1}$ h^{-1} using a standard curve with 4-methylumbelliferone added to the soil suspension of each sample.

4.2.3 Molecular analysis

4.2.3.1 DNA extraction from soil

Genomic DNA was extracted from 0.25 g frozen soil using PowerSoil DNA Isolation Kit® (MO BIO Laboratories, Inc., Carlsbad, CA) following the manufacturer's instructions, but with a slight modification in the homogenization and cell lysis step: TissueLyzer II (Qiagen, Valencia, CA) was used twice for 3 min at 30 Hz. DNA concentration and quality were assessed with a micro-volume UV/VIS spectrophotometer (UVS-99, UVIS-DROP; ACTGene, Inc., Piscataway, NJ).

4.2.3.2 Analysis of bacterial and fungal community structures using T-RFLP

Bacteria- and fungi-specific PCR was carried out in duplicates on the 16S rRNA and the ITS gene, respectively, using the primers and thermal profiles shown in Table 4.2. Reactions were carried out in 50 µL with 1x Taq buffer, 0.6 U μL^{-1} GoTaq DNA Polymerase (Promega, Madison, WI), 0.2 mM dNTP, 3 mM $MgCl_2$, 0.2 µM of each primer, 4 µL of genomic DNA and molecular grade water in a Labcycler (Sensoquest, Göttingen, Germany).

Bacterial 16S rDNA and fungal ITS amplicons were digested for 4 hours at 37°C with the restriction endonuclease *Msp*I (Promega). The reaction mixture contained: 1x restriction enzyme buffer, 2 µg acetylated bovine serum albumin, 5 U of restriction enzyme, 10 µL template and molecular grade water to a final volume of 20 µL. The digestion reaction was inactivated by incubation at 65°C for 15 min and the samples were purified with Sephadex G-50 (GE Healthcare, Glattbrugg, Switzerland), according to the manufacturer's instructions.

Between 1 and 4 µL of purified, digested PCR product was added to 10 µL HI-DITM Formamide (Applied Biosystems, Foster City, CA) and GeneScanTM-500 LIZ as internal standard (Applied Biosystems). Samples were heated at 95°C for 5 min, immediately chilled on ice and loaded into an ABI 3130XL genetic analyzer (Applied Biosystems), using POP 7 polymer (Applied Biosystems) as running polymer. The electropherograms were analyzed with GeneMapper Software 3.7 (Applied Biosystems), retaining T-RFs with fragment sizes between 50 and 500 bp. Runs with total peak areas between 100,000 and 400,000 fluorescent units were accepted for analysis. Duplicate runs of each sample were processed and analyzed with the T-REX software (Culman et al. 2009) as described in Mészáros et al. (2013). A data matrix based on the average peak areas of the duplicates was calculated.

Table 4.2. Primers and thermal profiles of PCR amplification of the community analysis.

Community	Targeted gene	Primer name	Primer sequence	Reference	Thermal profile	Analysis
Bacteria	16S rDNA	27F-FAM[1]	5'-AGA GTT TGA TCM TGG CTC AG-3'	Lane (1991)	95°C (5 min), followed by 35 cycles of 95°C (30 s), 56°C (30 s), and 72°C (1 min), completed by 10 min at 72°C	T-RFLP
		1406R	5'-GAC GGG CGG TGT GTR CA-3'	Baker et al. (2003)		
Fungi	ITS	ITS1F-FAM[1]	5'-CTT GGT CAT TTA GAG GAA GTA A-3'	Lord et al. (2002)	95°C (5 min), followed by 35 cycles of 95°C (30 s), 53°C (30 s), and 72°C (1 min), completed by 10 min at 72°C	T-RFLP
		ITS4R	5'-TCC TCC GCT TAT TGA TAT GC-3'	Lord et al. (2002)		
phoD	*phoD*	*phoD*-F733	5'-TGG GAY GAT CAY GAR GT-3'	Ragot et al. (2015)	95°C (5min), followed by 35 cycles of 95°C (30s), 58°C (30s) and 72°C (30s), completed by 10 min at 72°C	454-sequencing
		phoD-R1083	5'- CTG SGC SAK SAC RTT CCA-3'	Ragot et al. (2015)		
phoX	*phoX*	*phoX*-R455	5'- CAG TTC GGB TWC AAC AAC GA-3'	This study	95°C (5min), followed by 35 cycles of 95°C (30s), 60°C (30s) and 72°C (30s), completed by 10 min at 72°C	454-sequencing
		phoX-R1076	5'- CGG CCC AGS GCR GTG YGY TT -3'	This study		

[1]Labeled with 6-carboxyfluorescein (FAM)

4.2.3.3 Analysis of *phoD*- and *phoX*-harboring community structure and composition using 454-sequencing

Prior to sequencing, DNA extracts were tested for presence of *phoD* and *phoX* by PCR using the primers and thermal profiles shown in Table 4.2. Duplicate DNA extracts were then pooled and adjusted to a concentration of 20 ng μL^{-1}. Pooled extracts were sequenced on a GS-FLX+ platform (Roche 454 Life Sciences, Branford, CT) by Research and Testing Laboratory (Lubbock, TX, http://www.researchandtesting.com). Barcoded primer design, sequencing, and initial quality filtering were carried out by Research and Testing Laboratory (RTL, Lubbock, TX). Briefly, sequences with a quality score below 25 were trimmed. Chimeras were removed using USEARCH with a clustering at a 4% divergence (Edgar et al. 2011). Denoising was carried out using the Research and Testing Denoiser algorithm, which uses non-chimeric sequences and the quality scores to create consensus clusters from aligned sequences. For each cluster, the probability of prevalence of each nucleotide was calculated and accordingly, a new quality score was generated, which was then used to remove noise from the dataset.

4.2.4 Sequence analysis

Sequencing datasets of *phoD* and *phoX* were analyzed separately using MOTHUR (Schloss et al. 2009). Sequences were analyzed as nucleic acid sequences to keep the maximum information and allow accurate identification. After demultiplexing, reads including ambiguities, mismatches with the specific primers and the barcode were removed. Reads with an average quality score below 20 were then filtered. Resulting reads were trimmed at 150 bp and 400 bp for *phoD* and at 250 bp and 600 bp for *phoX* as minimum and maximum length, respectively. Sequences were aligned using the Needleman-Wunsch global alignment algorithm as implemented in MOTHUR, using 6-mers searching and the aligned reference sequences as template. The pairwise dis-

tance matrix was calculated from the alignment and sequences were clustered using the k-furthest method as implemented in MOTHUR. A similarity cutoff at 75% and 57% was used to define the operational taxonomic units (OTUs) of *phoD* (Tan et al. 2013) and of *phoX* (See Chapter 3), respectively. OTU matrices were then normalized to the smallest library size using the normalized.shared command in MOTHUR to allow comparison between samples, by dividing the relative abundance of each OTU by the total number of reads per sample. The normalized values were then rounded to the nearest integer (Table S4.3).

Taxonomy assignment was carried out using blastn in BLAST+ (Camacho et al. 2009) with a minimum e-value of 1e-8 to retrieve NCBI sequence identifiers (GI accession number). Subsequently, in-house Perl scripts were used to populate and query a mysql database containing the NCBI GI number and taxonomic lineage information (script written by Stefan Zoller, Genetic Diversity Centre, ETH Zurich, available on request) (Tables S4.4 and S4.5).

4.2.5 Data analysis

Correlations between environmental factors were evaluated using Pearson's linear regression in **R** v.2.15.0 (R Core Team, httpt://www.R-project.org/) (Table S4.1). ANOVA was used to test for significant effects of geographical origin and land-use on environmental factors, also in **R** .

Significant differences between the community structures were tested using analysis of dissimilarity (*anosim*) for the bacterial and fungal communities and pairwise *Libshuff* analysis as implemented in MOTHUR with 1000 iterations (Schloss et al. 2004) for the *phoD*- and *phoX*-harboring communities. Chao1 species richness and evenness indices were calculated using PAST software (Hammer et al. 2001) for the bacterial and fungal communities and using EstimateS (Version 9, httpt://purl.oclc.org/estimates) for the *phoD*- and *phoX*-harboring communities (Table S4.1). Rarefaction curves were

calculated and extrapolated to 5000 and 3000 reads for *phoD*- and *phoX*-harboring communities, respectively, to standardize the samples using EstimateS (Figure S4.3).

Non-metric multidimensional scaling analysis (NMDS) was carried out to analyze the variation within the bacterial, fungal, *phoD*- and *phoX*-harboring communities, using the vegan package (vegan: Community Ecology Package. **R** package version 2.2-0. http://CRAN.R-project.org/package=vegan) in **R**. Redundancy analysis (RDA) was carried out to evaluate the correlation between environmental factors and community structures in **R** (Figures S4.1 and S4.2). Prior to analysis, the environmental factors were standardized using z-scores (Ramette 2007).

4.3 Results

4.3.1 Soil properties in the two geographical origins, as affected by land-use

The samples displayed a large range of soil properties. For example, pH ranged from 4.2 to 8.3 and TOC from 3.7 to 59.6 g kg^{-1} soil (Table 4.3). A 38-fold difference in P_{res} and P_{org} was observed between the samples, while TP varied by a factor of 10.

Australian soils had significantly lower concentrations of TOC, TN, TP, P_{org}, C_{mic} and N_{mic} than Swiss soils (Table 4.3). Potential alkaline phosphatase activity was on average 10-fold higher in the Swiss than in the Australian soils, while potential acid phosphatase activity was similar for both geographical origins.

Land-use affected many soil properties (Table 4.3). Soil pH was generally lower in forest soils than in arable and grassland soils. The lowest TOC values were measured in arable soils and the highest in forest soils. Arable soils had the highest TP, P_{res} and P_{org} and the lowest C_{mic} and N_{mic}. The highest potential acid phosphatase activity was found in forest soils, and the highest potential alkaline phosphatase activity in grassland soils.

Table 4.3. General soil physicochemical and biological properties (mean of three analytical replicates). Letters indicate significant (<0.05) differences observed between samples from different geographical origins and from different land-uses.

Sample		Texture	pH	$CaCO_3$-C	TOC C	TN	TP	P_{res}	P_{org}	C_{mic}	N_{mic}	Potential acid phosphatase activity	Potential alkaline phosphatase activity
		(clay, silt, sand) g kg^{-1}		g kg^{-1} soil			mg kg^{-1} soil					nmol substrate g^{-1} soil h^{-1}	
AUS1-A		185, 333, 482	5.6	0	59.6	3.9	927	101.5	232	222	4.3	7928	117
AUS1-F		177, 278, 545	5.5	0	52.6	2	470	8.2	63	281	25.7	3055	92
AUS1-Gf		130, 252, 618	4.8	0	22.8	1.3	261	7.7	45	97	12.4	2574	36
AUS1-G		128, 247, 625	4.9	0	19.7	1.1	229	3	61	156	15.6	2602	24
AUS2-F		136, 461, 403	5.3	0	22.8	0.9	167	2.7	44	154	16.7	2914	41
AUS2-Gf		138, 247, 616	5	0	23.2	1.3	234	5.2	66	114	15.2	2385	45
AUS2-G		138, 276, 586	4.9	0	21	1.2	211	2.5	131	118	15.1	2560	43
AUS3-A		336, 194, 469	6.3	0	7.8	0.4	259	54	24	58	2.5	490	94
AUS3-F		120, 119, 761	5.8	0	28.5	1	235	10.3	27	118	10.1	2135	52
AUS3-G		381, 270, 349	6.9	0	23.7	1.5	705	33.7	74	252	19.4	2064	720
AUS4-A		265, 315, 419	5.4	0	16.3	1.1	616	53	58	86	5.7	1755	45
AUS4-F		274, 337, 389	6.4	0	20.1	1	500	24.3	25	187	11.7	1242	163
AUS4-G		298, 334, 368	5.7	0	15	0.9	466	6.5	56	140	12.8	1352	22
AUS5-F		108, 53, 839	7.0	0	3.7	0.2	167	14.5	10	54	8	88	72
AUS5-G		119, 113, 768	8.3	0	5.0	0.3	193	8.8	16	83	11.1	71	177
CH1-A		185, 680, 135	5.9	0	14.5	1.1	771	25.1	235	144	16.6	1027	121
CH1-F		487, 425, 88	5.5	0	45.2	2.8	599	16	228	487	53.9	2803	939
CH1-G		201, 694, 105	6.3	0	27.8	2.2	808	32.1	289	362	45	1418	30
CH2-A		242, 276, 482	7.4	14.4	39.9	3	1373	72.3	206	365	59.5	1069	3820
CH2-F		287, 332, 380	5.1	0	51.4	2.8	450	16	162	397	62.7	6145	186
CH2-G		238, 299, 463	7.8	14.4	44.7	3.1	1667	70.5	144	319	50.1	922	3061
CH3-A		419, 397, 184	7.0	0	30.1	2.4	1059	21.7	404	232	26.7	522	1426
CH3-F		198, 410, 392	4.2	0	27.7	1.4	403	4.8	188	128	11.8	2758	16
CH3-G		400, 452, 148	7.4	0	44.4	3.4	932	17.9	340	754	97.1	1480	5327
CH4-A		214, 366, 419	6	0	22	1.6	1000	88.5	177	190	26	1701	201
CH4-F		211, 440, 349	4.3	0	30.4	1.3	283	3.8	88	157	20.9	3107	21
CH4-G		302, 332, 366	6.6	0	36.1	3	700	12	254	506	68.5	2177	3668
CH5-A		172, 249, 579	6	0	11	0.8	551	33.4	133	100	11	598	71
CH5-F		177, 377, 447	4.4	0	32.3	1.7	301	9.2	128	208	22.5	5925	69
CH5-G		174, 260, 566	5.9	0	22.4	1.7	457	18.8	195	316	45	1895	436
Geographical origins[1]	AUS	196 ± 9, 254 ± 10, 550 ± 15	5.9 ± 0.9	NA[3]	22.2 ± 13.5[a]	1.2 ± 0.8[a]	376 ± 222[a]	22.4 ± 27	59 ± 57[a]	141 ± 66[a]	12.4 ± 5.8[a]	2214 ± 1794	116 ± 168[a]
	CH	260 ± 10, 399 ± 13, 341 ± 16	6.0 ± 1.1	NA	32.0 ± 11.3[b]	2.2 ± 0.8[b]	760 ± 384[b]	29 ± 25.4	211 ± 82[b]	311 ± 171[b]	41.2 ± 23.8[b]	2236 ± 1674	1293 ± 1711[b]
Land-use[2]	Arable	253 ± 8, 351 ± 14, 396 ± 14	6.2 ± 0.6[a]	NA	24.0 ± 14.1[a]	1.8 ± 1.1	820 ± 324[a]	56 ± 27[a]	183 ± 110[a]	175 ± 94[a]	19.0 ± 17.6[a]	1886 ± 2330[a]	737 ± 1243[ab]
	Forest	218 ± 11, 332 ± 10, 449 ± 18	5.4 ± 0.9[b]	NA	31.5 ± 14.3[b]	1.5 ± 0.8	358 ± 141[b]	11.0 ± 6[b]	96 ± 72[b]	217 ± 127[a]b	24.2 ± 17.9[ab]	3017 ± 1757[b]	165 ± 263[b]
	Grassland	221 ± 11, 315 ± 13, 465 ± 20	6.2 ± 1.2[a]	NA	25.5 ± 11.0[ab]	1.8 ± 0.9	572 ± 413[b]	18 ± 19[b]	139 ± 102[ab]	268 ± 193[b]	33.9 ± 26.4[b]	1792 ± 741[a]	1132 ± 1745[a]

Soil property abbreviations: $CaCO_3$-C: Carbonate content, TOC: total organic C, TP: total P, P_{res}: resin-extractable P, P_{org}: organic P, C_{mic}: microbial C biomass, N_{mic}: microbial N biomass.

[1] Significance differences between geographical origin were tested using a one-way ANOVA comparing samples from Australia and Switzerland

[2] Significance differences between land-uses were tested using a two-way ANOVA with land-use and site as factors (i.e.: land-use + site)

[3] NA: non-applicable

In our data set, TOC was correlated with TN, TP, P_{org}, P_{res}, C_{mic} and N_{mic} (Table S4.1). Additionally, potential acid phosphatase decreased with increasing soil pH, while potential alkaline phosphatase activities had the opposite trend (Table S4.1).

4.3.2 Correlations between the bacterial and fungal community structures and environmental factors

Bacterial and fungal community structures clustered strongly according to their geographical origin (Figure 4.1). The variation within the bacterial community structures was smaller in the Australian than in the Swiss soils, whereas the variation within the fungal community structure was similar for both geographical origins.

Environmental drivers of both bacterial and fungal community structures were climate, soil group, land-use, soil texture and all P forms (TP, P_{res} and P_{org}), with P_{org} being a common highly significant driver (Table 4.4). The bacterial community structure was additionally correlated to pH, while the fungal community structure was affected by TOC and TN.

Table 4.4. Correlation between the bacterial, fungal, *phoD*- and *phoX*-harboring community structure and environmental factors based on Redundancy analysis (RDA) (Figures S4.1 and S4.2).

	Bacteria	Fungi	*phoD*	*phoX*
Climate	**<0.001**	0.022	**<0.001**	**<0.001**
Soil group	0.004	0.015	0.045	0.008
Land-use	**<0.001**	0.004	**<0.001**	0.007
Clay	n.s.	0.047	n.s.	n.s.
Silt	0.034	0.007	n.s.	n.s.
pH	0.007	n.s.	n.s.	0.043
$CaCO_3$-C	n.s.	n.s.	n.s.	n.s.
TOC	n.s.	0.043	**<0.001**	0.020
TN	n.s.	0.002	**<0.001**	**<0.001**
TP	0.002	0.006	0.016	0.050
P_{org}	**<0.001**	**<0.001**	**<0.001**	**<0.001**
P_{res}	0.019	0.047	0.013	0.045

n.s.:non-significant

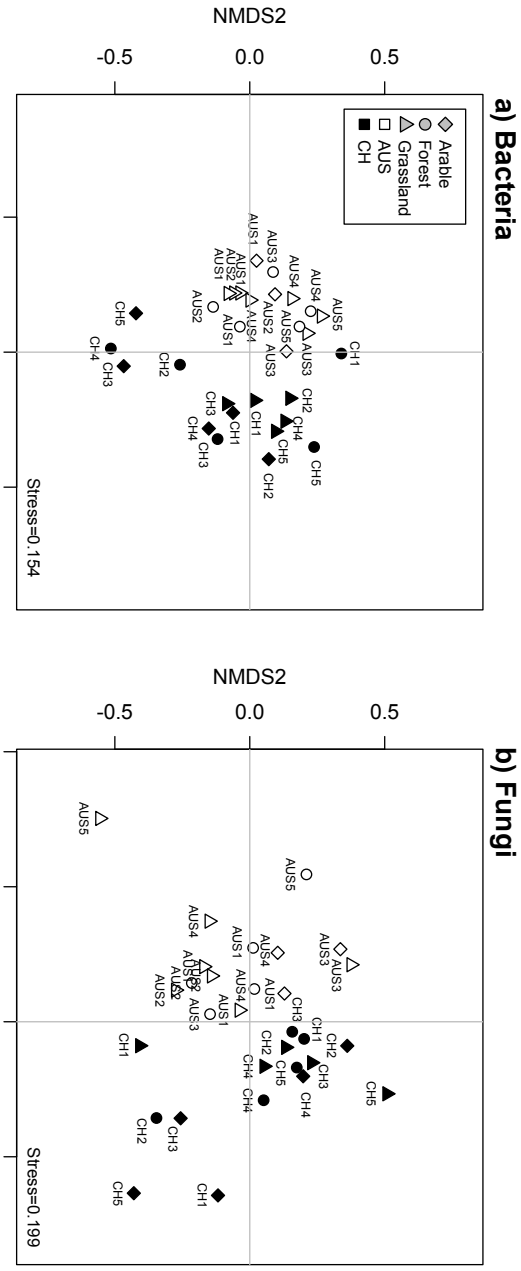


Figure 4.1. Non-metric multidimensional scaling (NMDS) analysis of the bacterial (a) and fungal (b) community structures.

4.3.3 Taxonomic composition, structure and main drivers of the *phoD*-harboring community

The *phoD* gene was found in 1 archaeal, 13 bacterial and 2 fungal phyla (Figure 4.2). The dominant *phoD*-harboring phyla which were found in all samples consisted of *Actinobacteria*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes* and *Proteobacteria*. *Acidobacteria*, *Bacteroidetes, Nitrospirae* and *Ascomycota* also represented dominant phyla in the *phoD*-harboring community, but only in some soil samples. No taxonomy could be assigned to < 1.0% of the reads in the samples. An additional 0.2 to 5.5% of the reads were highly similar to bacterial sequences from environmental samples but could not be assigned to a phylum.

The total *phoD*-harboring community structure was more variable within the Australian soils than within the Swiss soils, and the community was not separated according to geographical origins (Figure 4.4a).). The environmental drivers shaping the total *phoD*-harboring community structure were climate, soil group, land-use and all measured nutrient concentrations (TOC, TN, TP, P_{org}, P_{res}) (Table 4.4). Furthermore, the environmental factors affected the relative abundance of each *phoD*-harboring phylum differently. The most influential environmental factor affecting the relative abundance of *phoD*-harboring phyla was P_{res}, influencing the relative abundance of 11 of the 16 classified *phoD*-harboring phyla, including all dominant phyla (Figure 4.2). In more detail, the relative abundance of *phoD*-harboring *Actinobacteria*, *Cyanobacteria* and *Firmicutes* increased to an optimum P_{res} between 25 and 75 mg P kg^{-1} soil, whereas the relative abundance of *phoD*-harboring *Proteobacteria* was the lowest when P_{res} in this P_{res} class (Figure 4.5a). The relative abundance of *phoD*-harboring *Planctomycetes* increased with P_{res} and P_{org} (Figure 4.5a and b). The second most influential environmental factor affecting the relative abundance of *phoD*-harboring phyla was soil group, influencing the relative abundance of 9 of the 16 *phoD*-harboring phyla (Figure 4.3). *phoD*-harboring *Proteobacteria* were the most abundant phylum in the sampled Planosols,

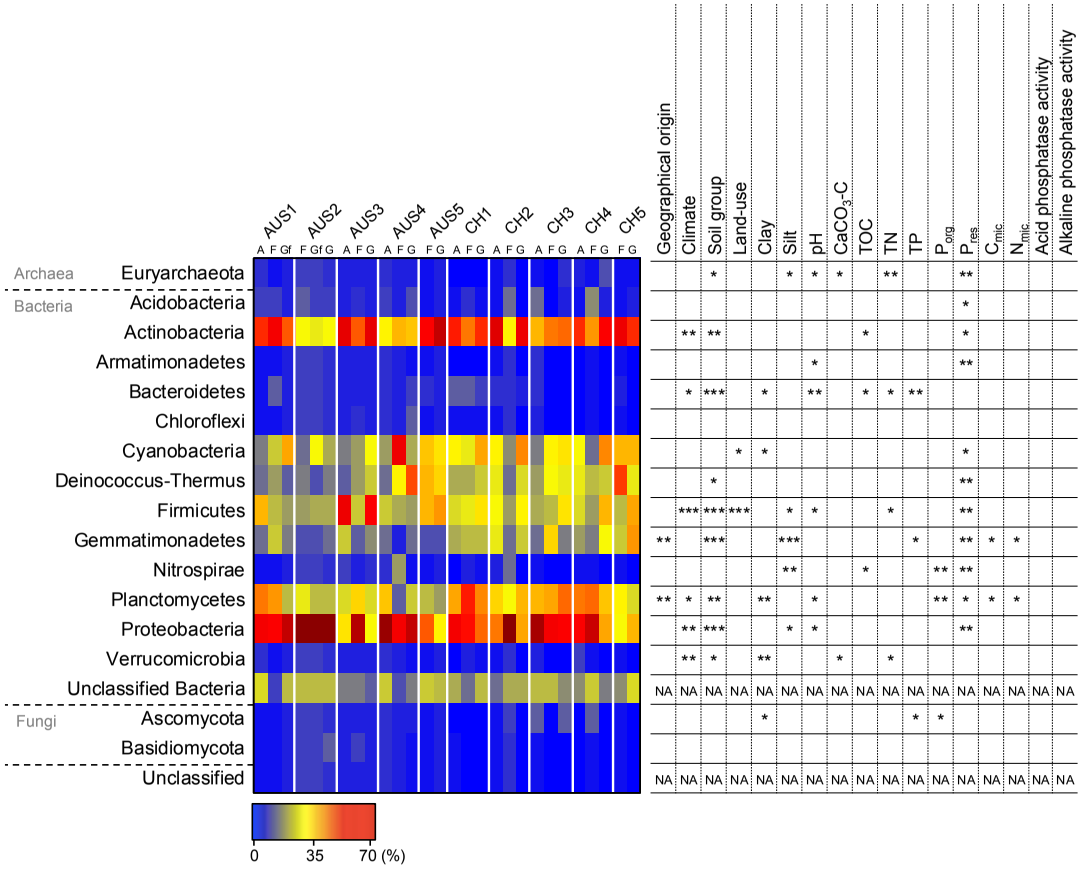


Figure 4.2. Relative abundance of *phoD*-harboring phyla and their correlation with environmental factors with *p*-values <0.05, <0.01 and <0.001 indicated by *, ** and ***, respectively.

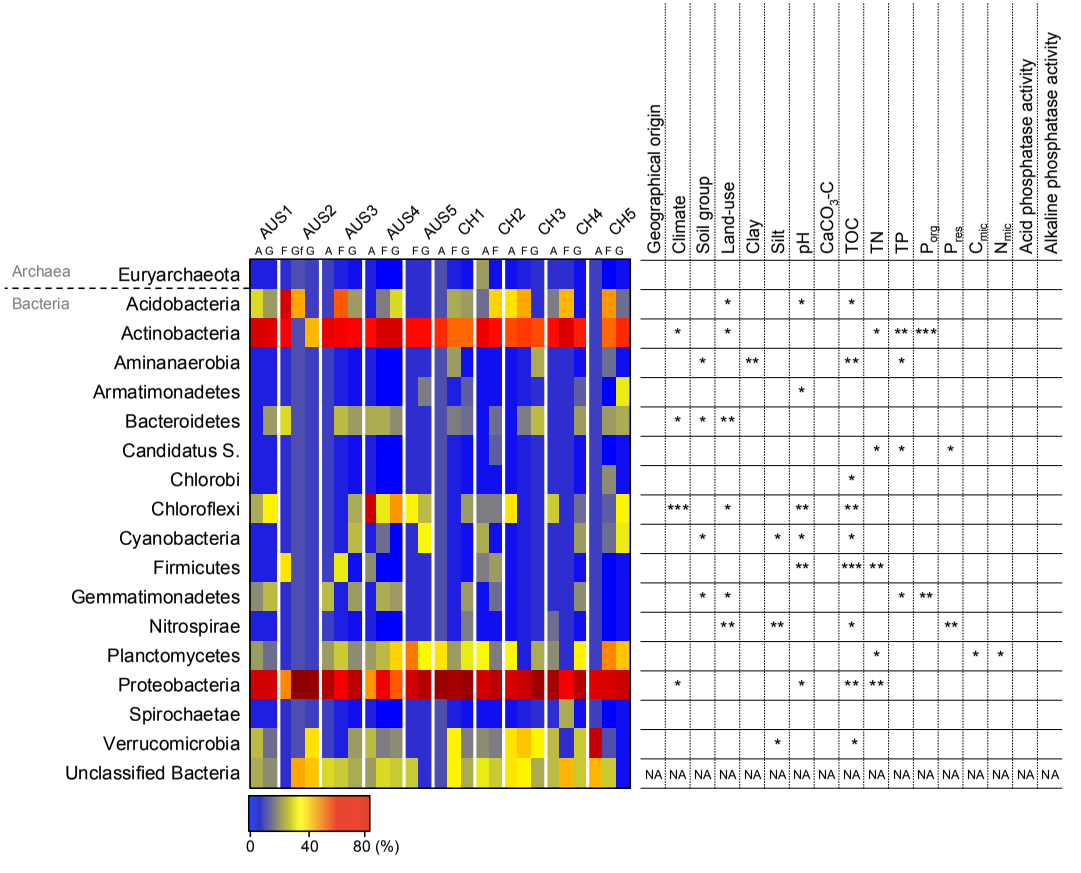


Figure 4.3. Relative abundance of *phoX*-harboring phyla and their correlation with environmental factors with *p*-values <0.05, <0.01 and <0.001 indicated by *, ** and ***, respectively.

Vertisols and Cambisols, while *phoD*-harboring *Actinobacteria* were predominant in the sampled Luvisols and Leptosols (Figure 4.5c). Land-use affected the relative abundance of *phoD*-harboring *Cyanobacteria*, which were more abundant in grassland than in arable and forest soils, and of *phoD*-harboring *Firmicutes*, which were significantly lower in forest than in arable and grassland soils (Figure 4.5a and d).

In summary, our results showed that climate, soil group, land-use and nutrient concentrations are major determinants of the total *phoD*-harboring community structure, and soil group and P_{res} are the principal environmental drivers governing the relative abundance of *phoD*-harboring phyla in the sampled soils. Additionally, despite the large range of environmental factors, the *phoD*-harboring composition was always made up of the same dominant phyla (*Actinobacteria*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Planctomycetes* and *Proteobacteria*).

4.3.4 Taxonomic composition, structure and main drivers of the *phoX*-harboring community

The *phoX* gene was found in 1 archaeal phylum and 16 bacterial phyla (Figure 4.3). *Proteobacteria* were the dominant *phoX*-harboring microbes in all soils, ranging between 15 and 72% of the total *phoX*-harboring community. However, the *phoX*-harboring phyla *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Planctomycetes* and *Verrucomicrobia* composed between 16 and 49% of *phoX*-harboring community in some soils, resulting in contrasting *phoX*-harboring community compositions between samples. Unclassified sequences represented between 3.9 and 5.0% of the reads in the samples and were all affiliated to the bacterial kingdom.

As for *phoD*, variation in the *phoX*-harboring community structure was generally higher in the Australian than in the Swiss soils (Figure 4.4b). Three Swiss soils (CH3-A, CH4-G and CH5-A), however, clustered away from the other soils, suggesting that their *phoX*-harboring community structure was significantly different from the other

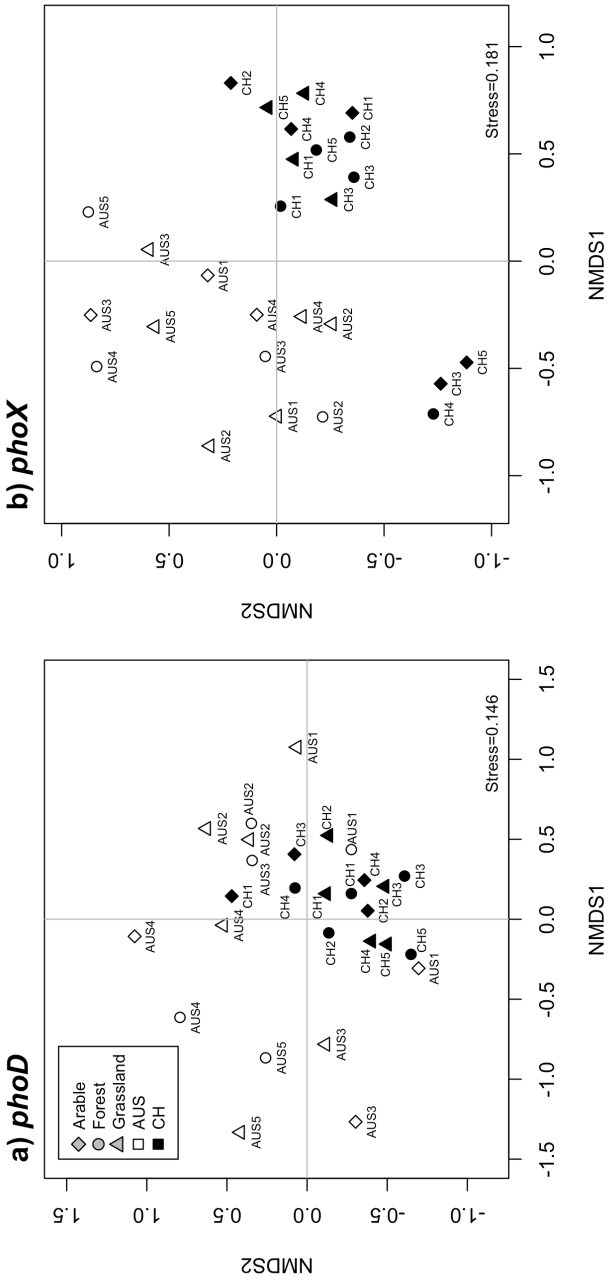


Figure 4.4. Non-metric multidimensional scaling (NMDS) analysis of the *phoD*- (a) and *phoX*- (b) harboring community structure.

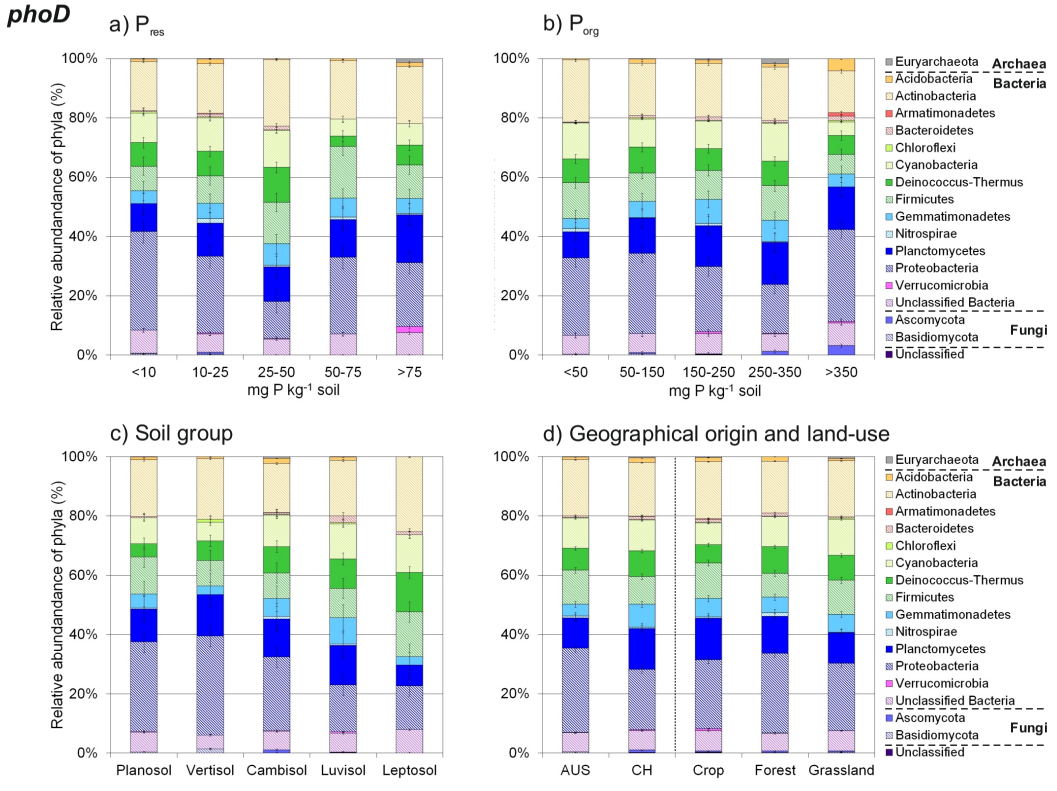


Figure 4.5. Average of the relative abundance of *phoD*-harboring phyla depending on P_{res} (a), P_{org} (b), soil group (c) and geographical origins and land-use (d).

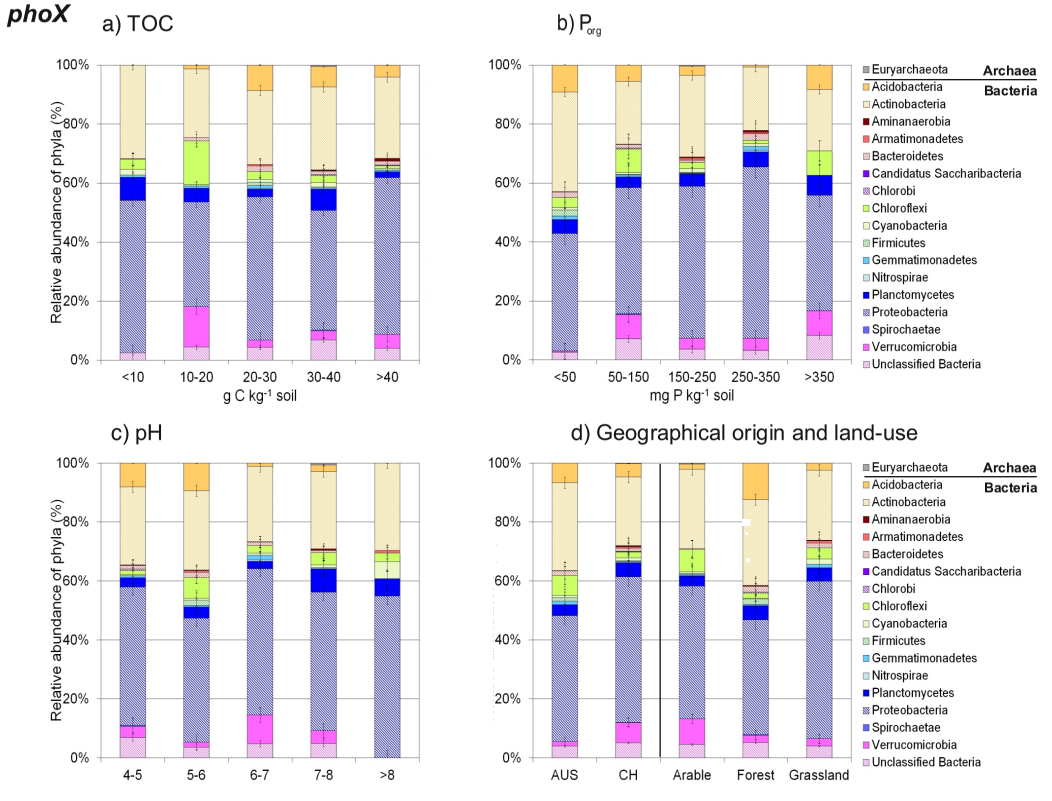


Figure 4.6. Average of the relative abundance of *phoX*-harboring phyla depending on TOC (a), P_{org} (b), pH (c) and geographical origins and land-use (d).

soils (Figure 4.4b). The significant environmental drivers of the total *phoX*-harboring community structure were climate, soil group, land-use, pH and all measured nutrient concentrations (TOC, TN, TP, P_{org}, P_{res}) (Table 4.4). At the phylum level, however, TOC was the most influential environmental driver, influencing the relative abundance of 7 out of 17 *phoX*-harboring phyla (Figure 4.3). In more detail, *phoX*-harboring *Acidobacteria* reached the highest relative abundance at a TOC of 20 to 30 g kg^{-1} soil (Figure 4.6a). The relative abundance of *phoX*-harboring *Chloroflexi* and *Verrucomicrobia* was substantially higher at a TOC of 10 to 20 g kg^{-1} soil. Furthermore, the relative abundance of *phoX*-harboring *Actinobacteria* was greatest when TOC was below 10 g kg^{-1} soil. P_{org} was significantly correlated with the relative abundance of *phoD*-harboring *Actinobacteria*, which decreased with P_{org}, and *phoD*-harboring *Gemmatimonadetes*, which reached the highest relative abundance at a P_{org} of 250 to 350 mg kg^{-1} soil (Figure 4.6b). Land-use and pH were the second most important environmental drivers, each influencing the relative abundance of 6 out of 17 *phoX*-harboring phyla (Figure 4.3). As land-use and pH were correlated (Table S4.1), the effect of pH could not be distinguished from that of land-use. For example, the relative abundance of *phoX*-harboring *Acidobacteria* was significantly higher in forest soils characterized by a lower pH than in arable and grassland soils (Figure 4.6c and d). Nonetheless, a few *phoX*-harboring phyla clearly reacted to either land-use or pH. The relative abundance of *phoX*-harboring *Proteobacteria* was highest in grassland soils, followed by arable and forest soils (Figure 4.6d). *phoX*-harboring *Chloroflexi* reached the highest relative abundances at an optimum pH of 5 to 6, and *phoX*-harboring *Verrucomicrobia* at an optimum pH of 6 to 7 (Figure 4.6c).

Our results showed that climate, soil group, land-use, pH and nutrient concentrations impact the total *phoX*-harboring community structure, and that TOC, land-use and pH are essential environmental drivers of the relative abundance of *phoX*-harboring phyla in the sampled soils. The dominant *phoX*-harboring phyla varied substantially between soils.

4.4 Discussion

4.4.1 Effect of land-use on soil physicochemical and biological properties

The soils sampled in this study covered a large range of pH and total C, N and P (TP, P_{org}, P_{res}) concentrations comparable to other global studies (Fierer and Jackson 2006; Lauber et al. 2008; Lauber et al. 2009; Griffiths et al. 2011). Our results showed that land-use affected several soil properties. Arable soils were characterized by a higher TP, P_{res}, P_{org} but lower C_{mic} and N_{mic} than forest and grassland soils (Table 4.3).

In accordance with our results, arable management has been shown to improve P content in soil, principally via fertilization (Guggenberger et al. 1996; Vogeler et al. 2009), and to decrease microbial biomass and biological activity (e.g. phosphatase activity), mainly through soil disturbance and N and P fertilization (Feller et al. 1994; Puget and Lal 2005; Wallenstein et al. 2006). Likewise, forest soils generally have higher TOC and are more acidic than arable and grassland soils (Guo and Gifford 2002; Balota et al. 2014). The relatively low P_{res} and acidic pH of the forest soils can explain the high potential acid phosphatase activity in our study. The high potential alkaline phosphatase activity in grassland soils is likely linked to the high microbial activity characteristic of grassland soils, which is stimulated by high root density, biomass and root exudates (Steenwerth et al. 2002). Thus, our results point to the strong correlation between land-use and soil properties, and correlations of potential acid phosphatase and alkaline phosphatase activity with soil pH and microbial activity in the sampled soils.

4.4.2 Main environmental drivers of the bacterial and fungal community structures

Our results across 30 soils and 3 land-uses showed that major environmental factors shaping the bacterial and fungal community structures were climate, soil group, land-

use and soil P concentration (TP, P_{res}, P_{org}), with P_{org} the most closely correlated with community structure (Table 4.4). The bacterial community structure was also affected by pH, whereas the fungal community structure was influenced by TOC and N content.

Our results support the findings of Drenovsky et al. (2004), that climate and land-use govern bacterial and fungal community structure and composition. Additionally, the main soil properties influencing the bacterial community have been shown to be pH (Lauber et al. 2009; Griffiths et al. 2011) and P availability (Siciliano et al. 2014), which are in turn impacted by land-use (Balota et al. 2014). The absence of pH effect on the bacterial community structure in our study may be due to the fact that no soil with a pH < 4 was sampled. The fungal community structure has been reported to be strongly affected by land-use and location (Birkhofer et al. 2012). Nutrient status, mainly N and P, has previously been shown to be an important determinant of the fungal community structure (Lauber et al. 2008; Bissett et al. 2011).

4.4.3 Main environmental drivers of *phoD*-harboring community structure and composition

The main environmental factors shaping the total *phoD*-harboring structure across the 30 soils and 3 land-uses sampled in this study were climate, soil group, land-use and all measured nutrient concentrations (TOC, TN, TP, P_{org}, P_{res}) (Table 4.4). At the phylum level, the relative abundances of the *phoD*-harboring groups were primarily affected by P_{res} and soil group (Figure 4.2, 4.6a and c).

Fertilization with manure, water-soluble fertilizer and compost in arable and grassland soils has been reported to change the *phoD*-harboring community structures (Chhabra et al. 2013; Fraser et al. 2015b; Sakurai et al. 2008). Both applications of manure and water-soluble P fertilizer can lead to higher P_{org} (Garg and Bahl 2008; Schefe et al. 2015). Fertilization with N (270 kg N ha^{-1} yr^{-1}) and P (240 kg P ha^{-1} yr^{-1}) but not P fertilization alone led to a change in the *phoD*-harboring community structure in Chilean Andosol

pastures (Jorquera et al. 2014). On the other hand, P fertilization significantly changed the *phoD*-harboring community structure in a long-term water-soluble P fertilization trial on pasture in Ireland (Tan et al. 2013). The same dominant groups *Actinobacteria*, *Cyanobacteria* and *Proteobacteria* composed the *phoD*-harboring community in all treatments, suggesting that the *phoD*-harboring community composition is rather stable at different P input intensities (low, medium, high). Tan et al. (2013) also observed that abundance of *phoD*-harboring *Acidobacteria* decreased with increasing P fertilization, supporting our results that the relative abundance of *phoD*-harboring *Acidobacteria* is negatively correlated with P_{res}. In our study, however, P_{res} influenced the relative abundances of additional *phoD*-harboring phyla such as *Actinobacteria*, *Cyanobacteria*, *Firmicutes*, *Proteobacteria* and *Planctomycetes*. These changes in relative abundance depending on P_{res} may be attributed to a selective process of *phoD*-harboring phyla that can efficiently produce PhoD and access P from organic compounds in soil under P starvation conditions (Rengel and Marschner 2005). Hence, P_{org} is also an important environmental factor for *phoD*-harboring community in our samples as shown by our results (Table 4.4).

In agreement with our study, organic C has been reported as an important driver of the *phoD*-harboring community structure in subtropical orchard soils (Cui et al. 2015). Although soil pH has also been reported to be a strong determinant of the *phoD*-harboring community structure in grassland and in cropping systems (Wang et al. 2012b; Ragot et al. 2015), in this study soil pH was not correlated with the total *phoD*-harboring community structure. This absence of a pH effect may be due to the greater variety of land-uses and soil groups included in this compared the other studies (Wang et al. 2012b; Ragot et al. 2015). Nevertheless, pH was an important driver of the relative abundances of *phoD*-harboring *Euryarchaeota, Armatimonadetes, Bacteroidetes, Firmicutes, Planctomycetes* and *Proteobacteria*, suggesting that soil pH is a major determinant of some *phoD*-harboring phyla. Our results are in agreement with the findings based on the IMG/M database of Ragot et al. (2015), who reported the presence of the *phoD* gene in Archaea,

Bacteria and Fungi genomes. Additionally, they amplified the *phoD* gene from 13 different phyla in 6 grassland soils using the same primers as in this study. In our study, the *phoD* gene was amplified from three additional phyla, namely *Armatimonadetes, Ascomycota* and *Basidiomycota*. This difference may be due to the larger variety of land-uses, soil groups and soil properties included in this study that is likely to harbor a greater diversity of microorganisms.

4.4.4 Main environmental drivers of *phoX*-harboring community structure and composition

Using newly-designed primers, we showed that the *phoX*-harboring community structure is mainly shaped by climate, soil group, land-use, pH and all measured nutrient concentrations (TOC, TN, TP, P_{org}, P_{res}). At the phylum level, TOC, followed by pH and land-use were the leading environmental drivers, impacting the relative abundance of most *phoX*-harboring phyla.

The only insights into the *phoX*-harboring community in the environment available so far come from aquatic ecosystems. In Lake Taihu (China), the *phoX*-harboring community composition remained similar along a trophic gradient going from meso- to hypereutrophic and was predominantly composed of *Alpha-* and *Beta-Proteobacteria*, followed by *Gamma-Proteobacteria* and *Cyanobacteria* (Dai et al. 2014). In marine bacteria, the *phoX* gene was also found in *Actinobacteria*, Bacteroidetes, *Chloroflexi* and *Lentisphaerae* (Luo et al. 2009; Sebastián and Ammerman 2009). In contrast to these aquatic studies, we found the *phoX* gene in archaea and in an additional 10 bacterial phyla including *Acidobacteria*, *Planctomycetes* and *Verrucomicrobia*, which were highly abundant in several of our soil samples. The discrepancies observed between the *phoX*-harboring community composition in terrestrial and aquatic ecosystems suggest that different key players harbor the *phoX* gene in different types of environments. However,

these discrepancies may be attributed to the different primers used in these studies, which were designed based on marine microbial sequences.

In aquatic bacteria, the *phoX*-harboring community structure has been shown to be correlated with total P concentration with the highest diversity in hypereutrophic and the highest *phoX* gene abundance in mesotrophic regions (Dai et al. 2014). Our results together with the outcomes of their study suggest that P concentrations and particularly P_{org} are important determinants of the *phoX*-harboring community and particularly affects specific *phoX*-harboring phyla in both terrestrial and aquatic ecosystems. In contrast to the *phoD*-harboring community, the *phoX*-harboring community structure and composition was also affected by soil pH in our study. In line with this, different pH optima have been reported for *phoX*-harboring organisms, e.g. pH 7.5 and 10 for *Campylobacter jejuni* and *Synechococcus* sp., respectively (Van Mourik et al. 2008; Kathuria and Martiny 2011).

4.4.5 Composition and structure of *phoD*- or *phoX*-harboring communities in relation to potential alkaline phosphatase activity

Neither the structures of the *phoD*- and *phoX*-harboring community nor the relative abundances of *phoD*- or *phoX*-harboring phyla were correlated with the potential alkaline phosphatase activity in our samples. Moreover, potential alkaline phosphatase activity did not correlate with soil P concentrations (TP, P_{org}, P_{res}), whereas the *phoD*- and *phoX*-harboring community structures were both correlated with soil P concentrations (TP, P_{org}, P_{res}) and P_{org} in particular. Our results contrast with the findings of Fraser et al. (2015b) who observed a negative correlation between potential alkaline phosphatase activity and available P. Additionally, they reported a positive correlation between the abundance of the *phoD* gene and potential alkaline phosphatase activity, but did not find a correlation between the abundance of the *phoD* transcript and potential alkaline phosphatase activity.

It has been suggested that there is a high degree of functional redundancy in ecosystems so that ecological functioning remains generally unaffected by the exact species

composition harboring functional genes (O'Donnell et al. 2005). Additionally, the potential alkaline phosphatase activity measured in the laboratory is only a proxy of the potential alkaline phosphatase activity under natural conditions (Rocca et al. 2015). The potential alkaline phosphatase activity includes the activity of stabilized enzymes that persist in soil (Quiquampoix and Mousain 2005), and thus, does not represent the alkaline phosphatase production by the current microbial community (Burns et al. 2013). Key *phoD*- and *phoX*-harboring microorganisms may differ in the regulation and expression of *phoD* and *phoX*, leading to an uneven contribution to alkaline phosphatase production in soil. Moreover, not only PhoD and PhoX but also PhoA and other phosphatases, which were not analyzed here due to the lack of genetic tools, may contribute to the potential alkaline phosphatase activity measured in our samples. Further analysis combining potential alkaline phosphatase activity assay and genomics/transcriptomics of all three alkaline phosphatase genes is required to understand the link between potential alkaline phosphatase activity and the alkaline phosphatase genes.

4.5 Occurrence of *phoD* and *phoX* in various environments

Based on our meta-analysis of *phoX* using the metagenomes available in the IMG/M database (Figure 2.1b and 3.2), *phoX* was found in fewer ecosystems and in high abundance in terrestrial and marine ecosystems. In contrast, the presence of *phoD* was reported in many different ecosystems such as air, fresh water and marine ecosystems, microbiome associated to plant and animals and extreme environments such as deep sea hydrothermal vents (Ragot et al. 2015). This suggests that *phoX* is less widespread into the environment than *phoD*.

4.6 Conclusion

Our results suggest that the *phoD* gene is spread across the archaeal, bacterial and fungal kingdom, while the *phoX* gene is found in archaea and bacteria only. The *phoD*-harboring community was always composed of the same dominant phyla, suggesting that the *phoD*-harboring community composition is rather stable across various environmental factors. In contrast, dominant *phoX*-harboring phyla varied significantly between soils. Our results suggest that P_{org} followed by climate, soil group, land-use and other soil nutrient concentrations are common environmental drivers shaping biogeographical patterns of bacterial and fungal community structures as well as the *phoD*- and *phoX*-harboring community structure and composition. In addition, soil pH was an important driver of the bacterial and the *phoX*-harboring community structure. *phoD*-harboring phyla were principally affected by soil group and P_{res}, while *phoX*-harboring phyla were principally affected by TOC, followed by pH and land-use. The structure of the *phoD*- and *phoX*-harboring community seems partly independent from that of the bacterial community.

4.7 Acknowledgements

The authors would like to thank Dr. Stefan Zoller for the Perl scripts for taxonomic analysis and the Genetic Diversity Center (Zürich, Switzerland) for technical assistance. We also thank Monika Màcsai for her help with soil sampling, Laurie Schönholzer for TOC, TN and microbial C/N analyses. We acknowledge Agroscope (Switzerland) and the University of Sydney (Australia) for access to the "DOK" (Bio-Dynamic, Bio-Organic, and Conventional) system comparison trial, the long-term fertilization trials in Watt (Switzerland), Glenroy, Kia-Ora and Narrabri (Australia). This work was supported by the Swiss National Science Foundation (SNF, project 140900) and the University of Sydney.

4.8 Supplementary material

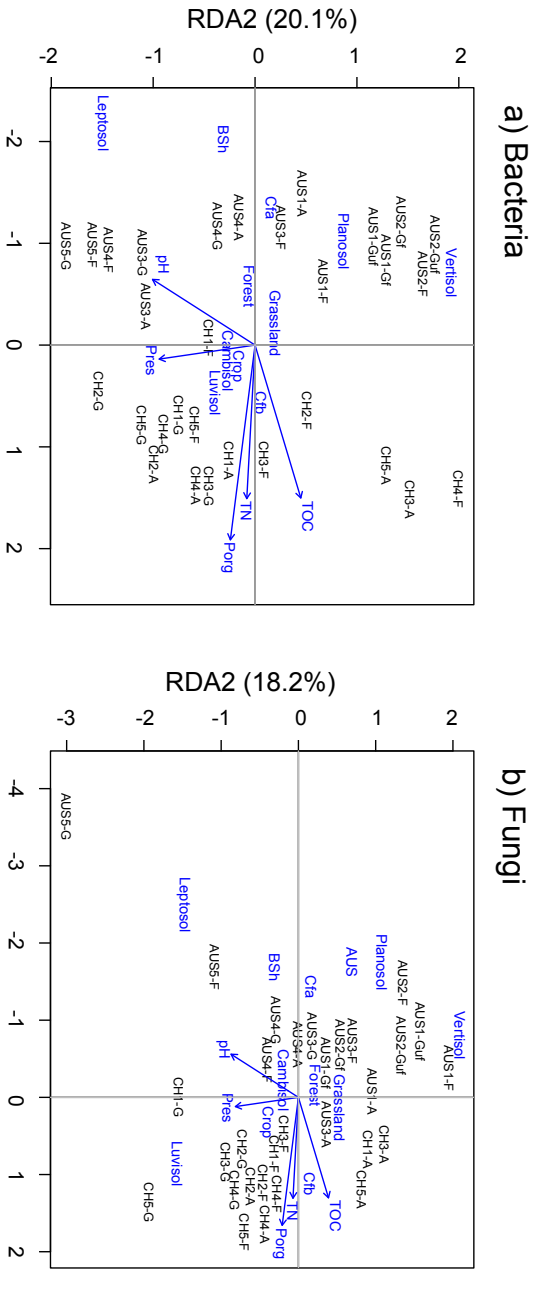


Figure S4.1. Redundancy analysis (RDA) of the bacterial (a) and fungal (b) community structures. Only significantly correlated environmental factors (continuous and categorical variables) are displayed.

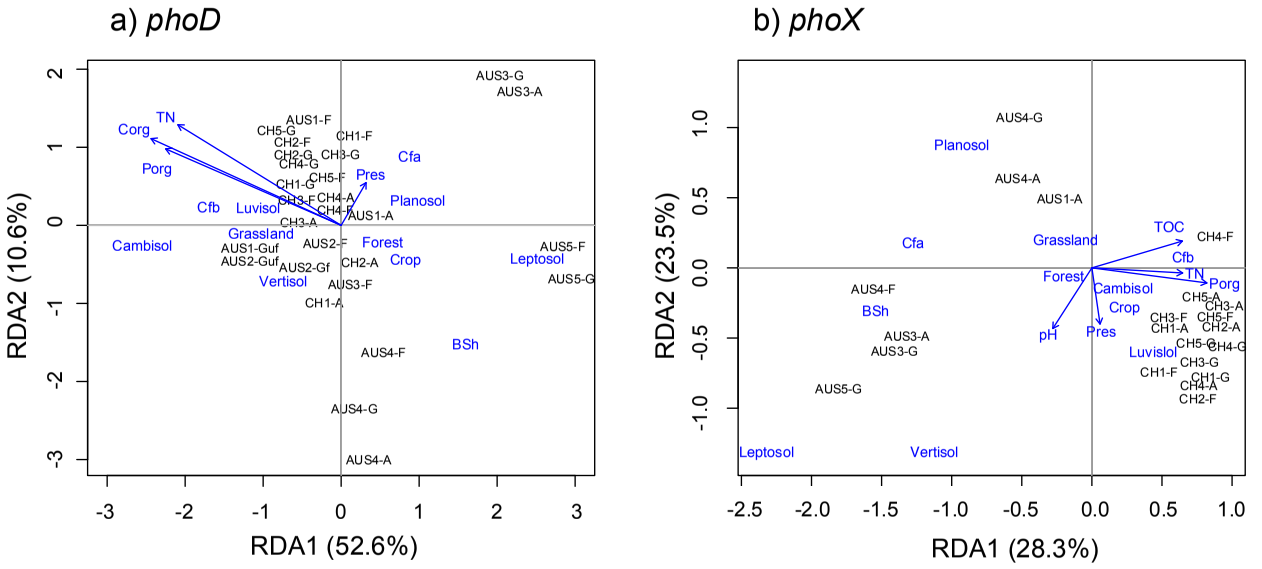


Figure S4.2. Redundancy analysis (RDA) of the *phoD*- (a) and *phoX*- (b) harboring community structures. Only significantly correlated environmental factors (continuous and categorical variables) are displayed.

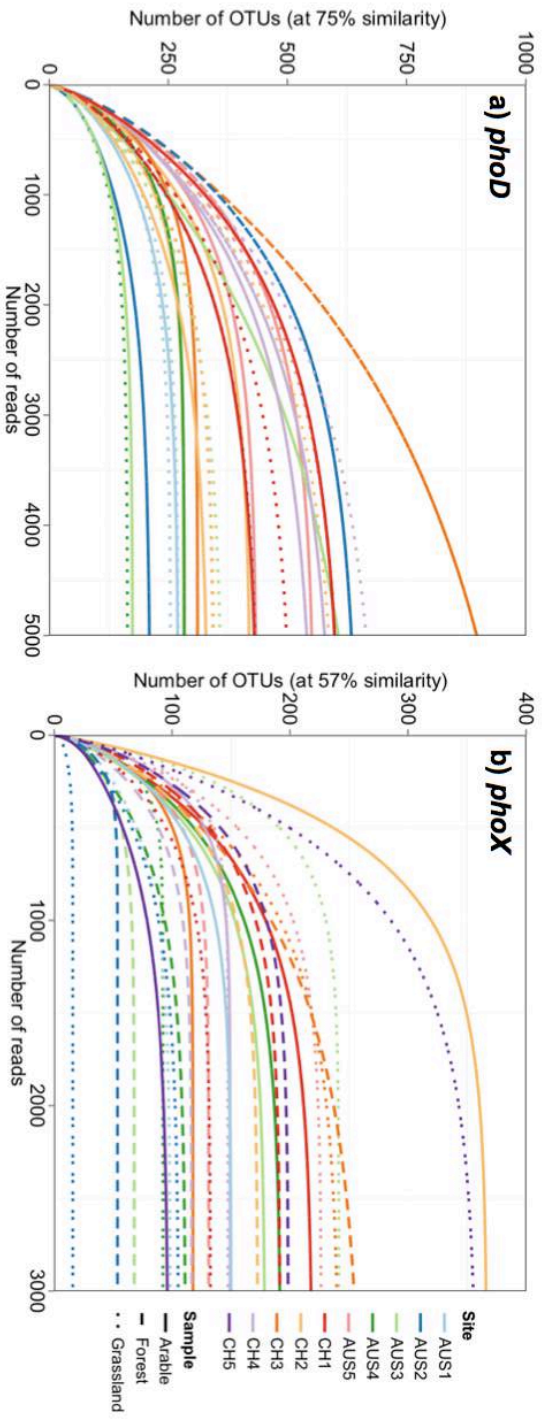


Figure S4.3. Rarefaction curves of phoD and phoX genes extrapolated to 5000 and 3000 reads, respectively.

Table S4.1. Correlation matrix of environmental variables across the entire dataset. Upper diagonal part contains correlation coefficient estimates and lower diagonal part contains corresponding *p*-values with p-values 0.001 indicated in bold.

	Geographical origin	Climate	Soil group	Land-use	Clay	Silt	pH	$CaCO_3$-C	TOC	TN	TP	P_{res}	P_{org}	C_{mic}	N_{mic}	Acid phosphatase activity	Alkaline phosphatase activity
Geographical origin	-	0.65	-0.57	n.s.	n.s.	0.53	n.s.	n.s.	0.37	0.5	0.52	n.s.	0.73	0.55	0.64	n.s.	0.436
Climate	**<0.001**	-	0.45	n.s.	n.s.	0.44	-0.33	n.s.	0.56	0.55	n.s.	n.s.	0.57	0.4	0.42	0.38	n.s
Soil group	**<0.001**	**<0.001**	-	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.46	n.s.	n.s.	n.s.	-0.38	n.s.	-0.41
Land-use	n.s.	n.s.	n.s.	-	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-0.52	n.s.	n.s.	n.s.	n.s.	n.s
Clay	n.s.	n.s.	n.s.	n.s.	-	n.s.	n.s.	n.s.	n.s.	0.45	0.43	n.s.	0.49	0.58	0.48	n.s.	0.463
Silt	0.003	0.015	n.s.	n.s.	n.s.	-	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.6	0.36	n.s.	n.s.	n.s
pH	n.s.	0.08	n.s.	n.s.	n.s.	n.s.	-	0.44	n.s.	n.s.	0.54	0.35	n.s.	n.s.	0.33	-0.58	0.565
$CaCO_3$-C	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.017	-	0.3	0.38	n.s.	0.46	0.1	0.2	0.33	-0.19	0.541
TOC	0.031	0.002	n.s.	n.s.	n.s.	n.s.	n.s.	0.004	-	0.88	0.47	0.19	0.48	0.7	0.62	0.6	0.445
TN	0.005	0.002	n.s.	n.s.	0.013	n.s.	n.s.	0.038	**<0.001**	-	0.7	0.41	0.72	0.8	0.72	n.s.	0.646
TP	0.003	n.s.	0.012	n.s.	0.019	n.s.	0.002	**<0.001**	**<0.001**	**<0.001**	-	0.7	0.62	0.5	0.5	-0.12	0.644
P_{res}	n.s.	n.s.	n.s.	0.003	n.s.	n.s.	0.055	0.011	n.s.	0.026	**<0.001**	-	n.s.	n.s.	n.s.	n.s.	n.s
P_{org}	**<0.001**	**<0.001**	n.s.	n.s.	0.006	**<0.001**	n.s.	n.s.	**<0.001**	**<0.001**	**<0.001**	n.s.	-	0.66	0.63	n.s.	0.533
C_{mic}	0.002	0.029	n.s.	n.s.	**<0.001**	0.048	n.s.	n.s.	**<0.001**	**<0.001**	0.005	n.s.	**<0.001**	-	0.95	n.s.	0.777
N_{mic}	**<0.001**	0.02	0.041	n.s.	0.007	n.s.	0.08	n.s.	**<0.001**	**<0.001**	0.005	n.s.	**<0.001**	**<0.001**	-	n.s.	0.81
acid phosphatase	n.s.	0.036	n.s.	n.s.	n.s.	n.s.	**<0.001**	n.s.	0.006	0.013	n.s.	n.s.	n.s.	n.s.	n.s.	-	n.s
alkaline phosphatase	0.016	n.s.	0.028	n.s.	0.01	n.s.	**<0.001**	0.002	0.003	0.012	0.017	n.s.	0.002	**<0.001**	**<0.001**	n.s.	-

Correlation matrix of environmental variables across the entire dataset. Upper diagonal part contains correlation coefficient estimates and lower diagonal part contains corresponding *p*-values with *p*-values < 0.001 indicated in bold.
n.s. : non-significant.

Table S4.2. Richness and evenness indices for the bacterial, fungal, *phoD*- and *phoX*-harboring community.

Sample		Bacteria		Fungi		*phoD*		*phoX*	
		Richness[1]	Evenness[2]	Richness	Evenness	Richness	Evenness	Richness	Evenness
AUS1-A		116	0.95	141	0.91	308	1	140	0.78
AUS1-F		119	0.95	124	0.88	1037	1	NA	NA
AUS1-GF		109	0.9	94	0.86	NA[5]	NA	88	0.69
AUS1-G		89	0.85	131	0.95	339	0.9	NA	NA
AUS2-F		120	0.90	127	0.95	265	0.75	119	0.80
AUS2-Gf		132	0.93	143	1.00	250	0.81	21.25	0.27
AUS2-G		109	0.93	139	1.00	266	0.85	213	0.67
AUS3-A		145	0.93	98	0.91	204	0.6	164	0.67
AUS3-F		113	0.93	132	0.95	640	1.00	113	0.71
AUS3-G		84	0.76	59	0.81	434	0.83	234	1.00
AUS4-A		85	0.83	132	0.95	676	0.81	174	0.80
AUS4-F		124	0.90	162	1.00	170	0.67	95.2	0.56
AUS4-G		94	0.85	122	1.00	355	0.81	116	0.84
AUS5-F		99	0.90	95	0.79	280	0.83	213	0.89
AUS5.G		116	0.90	71	0.79	159	0.67	125	0.78
CH1-A		113	0.88	59	0.60	429	0.96	241	0.71
CH1-F		63	0.71	126	0.95	549	0.98	179	0.87
CH1-G		110	0.85	119	0.86	606	0.90	119	0.76
CH2-A		126	0.88	116	0.93	414	0.88	345	0.98
CH2-F		128	0.93	108	0.91	280	0.90	162	0.78
CH2-G		109	0.88	185	1.00	594	0.90	NA	0.00
CH3-A		131	0.95	86	0.70	544	0.96	162	0.82
CH3-F		91	0.83	102	0.91	589	0.94	227	0.82
CH3-G		125	0.95	114	0.95	685	0.94	256	0.78
CH4-A		148	1.00	128	0.98	605	0.96	143	0.89
CH4-F		125	0.93	157	0.88	430	0.85	137	0.71
CH4-G		94	0.88	159	1.00	502	0.85	143	0.96
CH5-A		155	1.00	72	0.81	NA	NA	79	0.53
CH5-F		89	0.88	113	0.93	326	0.88	187	0.84
CH5-G		116	0.95	96	0.77	338	0.85	326	0.91
Geographical	AUS	110 ± 17	0.89 ± 0.05	110 ± 28	0.92 ± 0.07	385 ± 235	0.77 ± 0.24	140 ± 57	0.63 ± 0.29
Origins[3]	CH	115 ± 23	0.90 ± 0.07	116 ± 32	0.88 ± 0.11	492 ± 119	0.85 ± 0.23	193 ± 74	0.76 ± 0.23
Land-use[4]	Arable	127 ± 21^{a}	0.94 ± 0.05^{a}	104 ± 28	0.85 ± 0.13	454 ± 154	0.75 ± 0.32	148 ± 49	0.74 ± 0.15
	Forest	107 ± 20^{ab}	0.89 ± 0.07^{ab}	125 ± 21	0.92 ± 0.05	457 ± 244	0.88 ± 0.10	159 ± 43	0.70 ± 0.25
	Grassland	107 ± 14^{b}	0.89 ± 0.05^{b}	119 ± 34	0.92 ± 0.09	412 ± 159	0.78 ± 0.24	164 ± 86	0.64 ± 0.34

[1]Richness was calculated using the Chao1 index.
[2]Evenness was calculated using the J' index.
[3]One-way ANOVA comparing samples from Australia and Switzerland
[4]Two-way ANOVA with land use and site as factors comparing the land-uses including a site effect
[5]NA: Not-applicable

Table S4.3. Number of filtered reads (after initial processing), number of unique reads and number of reads after normalization per library for *phoD* and *phoX* genes.

Sample	*phoD*		*phoX*	
	Number of filtered reads	Number of normalized reads	Number of filtered reads	Number of normalized reads
AUS1-A	1671	624	3818	202
AUS1-F	1071	657	NA	NA
AUS1-GF	NA	NA	10344	214
AUS1-Guf	4751	629	NA	NA
AUS2-A	2888	632	1238	218
AUS2-Gf	6405	651	1107	213
AUS2-Guf	6357	637	1794	224
AUS3-A	4144	636	2617	200
AUS3-F	6858	638	1031	196
AUS3-G	2482	625	2819	191
AUS4-A	1008	611	2499	194
AUS4-F	1688	618	6159	194
AUS4-G	5782	613	1107	206
AUS5-F	1753	614	4999	192
AUS5.G	1386	632	3854	197
CH1-A	5755	583	1154	205
CH1-F	8625	586	2506	196
CH1-G	7261	597	8222	192
CH2-A	9867	597	3335	199
CH2-F	4150	586	7556	205
CH2-G	5680	598	NA	NA
CH3-A	8363	600	1195	198
CH3-F	5375	638	2585	203
CH3-G	5422	634	1264	209
CH4-A	6707	612	7255	198
CH4-F	11981	615	1471	197
CH4-G	4849	601	4776	199
CH5-A	NA	NA	6573	211
CH5-F	1629	647	4221	199
CH5-G	11318	653	10698	197

NA: Non-applicable

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples.

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
Total	1213	808	1145	1968	4608	5651	3741	4237	2189	925	311	6783	445	1093	3235	6494	6047	8434	2243	5121	5387	4270	4508	5122	5269	4197	1417	9930
Archaea	2	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1. Euryarchaeota	2	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1.1. Halomebacteria	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1.1.1. Haloarchaeales	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1.1.1.1. Haloarchaeaceae	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1.1.1.1.1. Haloferax	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	4	0	33	0	8
1.2. Methanomicrobia	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2.1. Methanomicrobiales	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2.1.1. Methanoregulaceae	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.2.1.1.1. Methanosphaerula	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bacteria	1144	806	1007	1916	4480	5465	3689	4188	2173	890	310	6664	424	1057	3079	6391	5794	8338	2184	4934	5186	4087	4392	4995	5006	4082	1391	9373
2. Acidobacteria	5	2	0	4	0	10	0	5	0	1	0	62	0	0	5	26	10	0	19	0	91	1	7	9	131	5	0	23
2.1. Acidobacteria	3	2	0	4	0	2	0	1	0	1	0	43	0	0	1	17	5	0	12	0	61	0	4	2	104	5	0	21
2.1.1. Acidobacteria subdivision 1	3	2	0	4	0	2	0	1	0	1	0	43	0	0	1	17	5	0	12	0	61	0	4	2	104	5	0	21
2.1.1.1. Acidobacteriaceae	3	2	0	4	0	2	0	1	0	1	0	43	0	0	1	17	5	0	12	0	61	0	4	2	104	5	0	21
2.1.1.1.1. Methanosphaerula	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.1.1.1.2. Chloracidobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
2.1.1.1.3. Acidobacter	3	2	0	4	0	2	0	1	0	1	0	39	0	0	1	17	5	0	12	0	59	0	4	1	59	0	0	12
2.1.1.1.4. Terriglobus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	45	5	0	0
2.2. Solibacteres	2	0	0	0	0	8	0	4	0	0	0	19	0	0	4	9	5	0	7	0	30	1	3	7	27	0	0	2
2.2.1. Solibacterales	2	0	0	0	0	8	0	4	0	0	0	19	0	0	4	9	5	0	7	0	30	1	3	7	27	0	0	2
2.2.1.1. Solibacteraceae	2	0	0	0	0	8	0	4	0	0	0	19	0	0	4	9	5	0	7	0	30	1	3	7	27	0	0	2
2.2.1.1.1. Solibacter	2	0	0	0	0	8	0	4	0	0	0	19	0	0	4	9	5	0	7	0	30	1	3	7	27	0	0	2
2.3. Actinobacteria	313	273	223	112	256	324	1515	924	869	65	38	750	140	556	1145	1157	1592	3444	154	1806	660	788	859	1319	770	1295	527	2591
2.3.1. Acidimicrobiales	0	0	0	0	0	0	0	0	0	1	0	2	0	0	1	6	0	0	0	0	0	0	0	0	0	0	0	0
2.3.1.1. Acidimicrobiaceae	0	0	0	0	0	0	0	0	0	1	0	2	0	0	1	6	0	0	0	0	0	0	0	0	0	0	0	0
2.3.1.1.1. Acidimicrobium	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
2.3.1.1.2. Ilumatobacter	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2.3.2. Actinomycetales	313	273	223	112	256	324	1515	924	869	64	38	748	140	556	1144	1151	1592	3444	154	1806	660	788	859	1319	770	1295	527	2591

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
2.3.2.1. Actinomycetaceae	171	69	159	28	77	184	1317	353	669	25	22	505	68	153	602	783	910	2047	36	1175	267	462	513	913	177	687	444	1623
2.3.2.1.1. Actinobaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
2.3.2.1.2. Actinomyces	30	7	10	1	0	2	629	48	18	1	0	137	7	10	159	106	113	328	6	194	16	37	117	103	9	154	7	231
2.3.2.1.3. Actinoplanes	4	3	0	0	9	64	404	30	308	1	12	9	5	0	73	36	64	129	2	97	17	21	31	34	12	79	20	70
2.3.2.1.4. Angiococcus	0	7	0	0	0	0	0	8	1	0	0	20	1	0	1	2	0	63	0	1	2	0	2	6	0	11	5	1
2.3.2.1.5. Kitasatospora	2	2	0	0	0	6	0	10	0	0	0	5	14	41	0	0	11	13	0	9	5	2	2	0	0	2	0	2
2.3.2.1.6. Micromonospora	11	1	11	0	0	10	25	13	30	0	0	5	5	7	24	130	61	303	2	131	11	19	53	67	6	9	58	79
2.3.2.1.7. Microstreptospora	28	15	7	1	4	13	131	26	90	4	0	35	5	3	13	20	24	100	0	46	15	15	16	8	7	19	26	323
2.3.2.1.8. Streptomyces	95	34	67	26	64	89	78	215	209	16	10	272	28	91	331	481	634	1088	24	638	178	367	289	688	134	409	327	908
2.3.2.1.9. Streptotrix	1	0	1	0	0	0	0	3	0	0	0	0	0	1	0	1	0	0	0	5	0	1	0	0	0	0	0	1
2.3.2.1.10. Thermomonospora	0	0	63	0	0	0	50	0	13	3	0	22	3	0	1	7	3	23	2	54	21	0	3	7	9	4	1	8
2.3.2.2. Actinosynnemataceae	27	5	2	3	1	0	22	74	60	2	5	27	22	323	24	22	37	44	21	30	17	4	13	54	18	28	6	121
2.3.2.2.1. Amycolatopsis	1	0	0	3	0	0	2	33	47	0	5	2	21	321	2	6	3	3	0	7	0	0	0	2	8	0	1	17
2.3.2.2.2. Kutzneria	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.3.2.2.3. Saccharothrix	26	5	2	0	1	0	14	19	13	2	0	17	1	2	21	16	34	41	21	20	17	4	10	52	5	28	5	104
2.3.2.2.4. Thermobispora	0	0	0	0	0	0	0	1	0	0	0	8	0	0	0	0	0	0	0	2	0	0	0	0	5	0	0	0
2.3.2.2.5. Thermopolyspora	0	0	0	0	0	0	6	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	3	0	0	0	0	0
2.3.2.3. Beutenbergiaceae	0	1	0	18	6	62	0	2	9	0	0	0	1	0	0	2	0	18	2	4	250	3	2	5	272	11	13	72
2.3.2.3.1. Beutenbergia	0	1	0	18	6	62	0	2	9	0	0	0	1	0	0	2	0	18	2	4	250	3	2	5	272	11	13	72
2.3.2.4. Catenulisporaceae	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
2.3.2.4.1. Actinomycetales	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
2.3.2.5. Corynebacteriaceae	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	1	0	16	0	4	0	0	7	0	1	6	0	5
2.3.2.5.1. Corynebacterium	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	1	0	16	0	4	0	0	7	0	1	6	0	5
2.3.2.6. Frankiaceae	6	7	37	0	38	8	6	102	0	5	11	91	6	22	15	54	8	54	6	30	16	17	11	39	47	11	0	8
2.3.2.6.1. Frankia	6	7	37	0	38	8	6	102	0	5	11	91	6	22	15	54	8	54	6	30	16	17	11	39	47	11	0	8
2.3.2.7. Geodermatophilaceae	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	1	0	0	0	0	0	0	2	0	3	0	0	0
2.3.2.7.1. Blastococcus	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
2.3.2.7.2. Geodermatophilus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0
2.3.2.7.3. Modestibacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
2.3.2.8. Glycomycetaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	19	2	0	9	23	5	0	1
2.3.2.8.1. Stackebrandtia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0	19	2	0	9	23	5	0	1
2.3.2.9. Gordoniaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	7	0	0	0	0	1	0	0
2.3.2.9.1. Gordonia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	7	0	0	0	0	1	0	0
2.3.2.10. Kineosporiaceae	12	8	0	0	0	0	0	0	0	0	0	0	0	0	18	74	18	138	0	83	6	0	9	0	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
2.3.2.10.1. Kineococcus	12	8	0	0	1	5	2	12	23	0	0	2	8	5	18	74	18	138	0	83	3	1	19	16	4	32	8	65
2.3.2.11. Micrococcaceae	22	10	0	0	17	8	11	51	15	2	0	16	2	6	16	9	45	123	22	32	21	3	14	21	1	50	1	70
2.3.2.11.1. Arthrobacter	0	0	0	0	2	3	5	20	0	0	0	12	1	0	8	2	12	99	15	10	2	0	9	6	1	42	0	61
2.3.2.11.2. Micrococcus	22	10	0	0	15	5	6	31	15	2	0	4	1	6	8	7	33	24	7	22	19	3	5	15	0	8	1	9
2.3.2.12. Microsphaeraceae	1	2	0	0	0	2	0	2	2	2	0	7	0	0	141	7	87	0	1	0	0	43	47	50	10	41	0	8
2.3.2.12.1. Nakamurella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3
2.3.2.12.2. Mycobacterium	1	2	0	0	0	2	0	2	2	2	0	7	0	0	141	7	87	0	1	0	0	43	47	48	10	41	0	5
2.3.2.13. Nocardiaceae	2	1	8	59	1	12	0	32	0	0	0	5	1	0	8	11	0	3	5	17	17	1	0	1	10	0	0	9
2.3.2.13.1. Nocardia	2	1	8	59	1	12	0	32	0	0	0	5	1	0	8	4	0	2	4	15	15	0	0	1	10	0	0	5
2.3.2.13.2. Rhodococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	1	1	2	2	1	0	0	0	0	0	4
2.3.2.14. Nocardioidaceae	68	170	17	4	116	48	157	308	114	28	0	91	40	52	311	184	480	993	61	431	28	253	223	227	208	455	63	674
2.3.2.14.1. Kribbella	67	55	17	4	116	48	157	300	114	28	0	69	40	51	311	184	475	980	61	426	28	253	208	227	207	454	63	667
2.3.2.14.2. Nocardiopsis	0	115	0	0	0	0	0	8	0	0	0	22	0	1	0	0	5	13	0	5	0	0	14	0	1	1	0	7
2.3.2.14.3. Thermobifida	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
2.3.2.15. Promicromonosporaceae	4	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	0	0	6	0	9	0	0	0	0	0
2.3.2.15.1. Isoptericola	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
2.3.2.15.2. Xylanimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
2.3.2.15.3. Microlunatus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
2.3.2.16. Tsukamurellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6	0	9	0	0	0	0	0
2.3.16.1. Tsukamurella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
2.3.3. Coriobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
2.3.3.1. Coriobacteriaceae	0	0	0	0	0	0	0	0	0	0	1	13	0	0	0	0	0	0	0	0	68	0	0	0	0	0	0	0
2.3.3.1.1. Adlercreutzia	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	68	0	0	0	0	0	0	0
2.3.3.1.2. Coriobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2.3.4. Rubrobacterales	0	0	0	13	0	41	4	0	26	0	4	18	54	12	1	1	0	8	0	9	0	0	0	4	4	4	0	7
2.3.4.1. Rubrobacteraceae	0	0	0	13	0	41	4	0	26	0	4	18	54	12	1	1	0	8	0	9	0	0	0	4	4	4	0	7
2.3.4.1.1. Rubrobacter	0	0	0	13	0	41	4	0	26	0	4	18	54	12	1	1	0	8	0	9	0	0	0	4	4	4	0	7
2.3.5. Solirubrobacterales	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2.3.5.1. Conexibacteraceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2.3.5.1.1. Conexibacter	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
3. Armatimonadetes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
3.1. Fimbriimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
3.1.1. Fimbriimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
3.1.1.1. Fimbriimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
3.1.1.1.1. Fimbriimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
4. Bacteroidetes	0	6	0	0	0	0	0	0	0	0	0	21	1	0	51	73	27	14	0	2	7	0	0	0	0	0	1	0
4.1. BCF group	0	6	0	0	0	0	0	0	0	0	0	21	0	0	51	73	27	14	0	2	7	0	0	0	0	0	1	0
4.1.1. BCF group	0	3	0	0	0	0	0	0	0	0	0	0	0	0	25	9	13	5	0	1	0	0	0	0	0	0	0	0
4.1.1.1. Rikenellaceae	0	3	0	0	0	0	0	0	0	0	0	0	0	0	25	9	13	5	0	1	0	0	0	0	0	0	0	0
4.1.1.1.1. Alistipes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
4.1.2. Cytophagales	0	0	0	0	0	0	0	0	0	0	0	19	0	0	1	51	1	4	0	0	7	0	0	0	0	0	0	0
4.1.2.1. Cytophaga-Flexibacter	0	0	0	0	0	0	0	0	0	0	0	19	0	0	1	51	1	4	0	0	7	0	0	0	0	0	0	0
4.1.2.1.1. Dyadobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	6	0	0	0	0	2	0	0	0	0	0	0	0
4.1.2.1.2. Fibrella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
4.1.2.1.3. Hymenobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	1	0	0	0	5	0	0	0	0	0	0	0
4.1.2.1.4. Spirosoma	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.3. Flavobacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
4.1.3.1. Flavobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
4.1.3.1.1. Flavobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4.1.3.1.2. Robiginitalea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
4.1.4. Sphingobacteriales	0	3	0	0	0	0	0	0	0	0	0	2	0	0	25	12	13	5	0	1	0	0	0	0	0	0	0	0
4.1.4.1. Chitinophagaceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4.1.4.1.1. Niabella	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.4.1.2. Niastella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
4.1.4.2. Saprospiraceae	0	3	0	0	0	0	0	0	0	0	0	0	0	0	25	9	13	5	0	1	0	0	0	0	0	0	0	0
4.1.4.2.1. Haliscomenobacter	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	0	7	3	0	5	3	0	0	0	0
4.2. Chlorobi	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.2.1. Chlorobiales	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.2.1.1. Chlorobiacea	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.2.1.1.1. Chlorobium	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Verrucomicrobia	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	2	0	0	0	1	0	0	24	0	0	0	0
5.1. Opitutales	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	1	0	0	24	0	0	0	0
5.1.1. Opitutaceae	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	1	0	0	24	0	0	0	0
5.1.1.1. Opitutus	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	1	0	0	24	0	0	0	0
5.1.2. Verrucomicrobiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
5.1.2.1. Verrucomicrobiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
5.1.2.1.1. Akkermansia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
6. Chloroflexi	0	0	0	0	0	0	0	2	0	0	0	50	0	0	2	2	2	0	0	0	1	0	0	0	0	0	0
6.1. Anaerolineae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0
6.1.1. Anaerolinaeles	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0
6.1.1.1. Anaerolinaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0
6.1.1.1.1. Anaerolinea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0
6.2. Caldilineae	0	0	0	0	0	0	0	2	0	0	0	36	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
6.2.1. Caldilineales	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6.2.1.1. Caldilineaceae	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6.2.1.1.1. Caldilinea	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6.2.2. Chloroflexaceaegroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.2.2.1. Roseiflexaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.2.2.1.1. Roseiflexus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.2.3. Herpetosiphonales	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.2.3.1. Herpetosiphonaceae	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.2.3.1.1. Herpetosiphon	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3. Thermomicrobia	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.3.1. Sphaerobacterales	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.3.1.1. Sphaerobacteraceae	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.3.1.1.1. Sphaerobacter	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
7. Cyanobacteria	18	39	149	10	263	132	55	88	159	17	120	208	51	100	358	435	809	695	35	858	71	339	423	357	64	673	172	1325
7.1. Cyanobacteria	0	0	0	0	0	1	0	0	0	0	0	33	0	0	10	1	0	0	3	0	0	0	2	0	0	1	0	57
7.1.1. Pleurocapsales	0	0	0	0	0	0	1	0	0	0	0	33	0	0	10	1	0	0	3	0	0	0	2	0	0	1	0	57
7.1.1.1. Chroococcidiopsis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	10	0	0	0	3	0	0	0	2	0	0	1	0	57
7.1.1.1.1. Chroococcidiopsis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	10	0	0	0	3	0	0	0	2	0	0	1	0	57
7.1.1.2. Pleurocapsa	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
7.1.1.2.1. Pleurocapsa	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
7.2. Gloeobacteria	16	36	149	10	241	103	53	79	150	14	120	45	51	92	294	266	722	580	26	804	40	167	384	257	63	621	168	1193
7.2.1. Gloeobacterales	16	36	149	10	241	103	53	79	150	14	120	45	51	92	294	266	722	580	26	804	40	167	384	257	63	621	168	1193
7.2.1.1. Gloeobacter	16	36	149	10	241	103	53	79	150	14	120	45	51	92	294	266	722	580	26	804	40	167	384	257	63	621	168	1193
7.2.1.1.1. Gloeobacter	16	36	149	10	241	103	53	79	150	14	120	45	51	92	294	266	722	580	26	804	40	167	384	257	63	621	168	1193
7.3. Nostocales	0	3	0	0	0	29	0	9	6	3	0	119	0	6	28	11	21	51	6	16	6	35	19	27	1	25	0	32
7.3.1. Nostocaceae	0	3	0	0	0	29	0	9	6	3	0	119	0	6	28	11	21	51	6	16	6	35	19	27	1	25	0	32

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
7.3.1.1. Amorphonostoc	0	0	0	0	0	29	0	9	6	3	0	119	0	5	3	2	8	46	6	15	5	0	0	16	1	24	0	32
7.3.1.1.1. Nostoc	0	0	0	0	0	29	0	9	6	3	0	119	0	5	3	2	8	46	6	15	5	0	0	16	1	24	0	32
7.3.1.2. Anabaena	0	3	0	0	0	0	0	0	0	0	0	0	0	1	25	9	13	5	0	1	1	35	19	11	0	1	0	0
7.3.1.2.1. Anabaena	0	3	0	0	0	0	0	0	0	0	0	0	0	1	25	9	13	5	0	1	1	35	19	11	0	1	0	0
7.4. Oscillatoriophycideae	2	0	0	0	22	0	1	0	3	0	0	11	0	2	26	157	66	64	0	38	25	137	18	73	0	26	4	43
7.4.1. Chroococcales	2	0	0	0	22	0	1	0	3	0	0	11	0	2	26	157	66	64	0	38	25	137	18	73	0	26	4	43
7.4.1.1. Chroococcus	2	0	0	0	22	0	1	0	3	0	0	11	0	2	26	157	66	64	0	38	25	137	18	73	0	26	4	43
7.4.1.1.1. Chroococcus	2	0	0	0	22	0	1	0	2	0	0	11	0	2	26	157	66	62	0	38	25	137	18	73	0	26	4	43
7.4.1.1.2. Cyanothece	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
7.4.2. Oscillatoriales	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.2.1. Microcoleus	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.2.1.1. Microcoleus	0	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8. Deinococcus-Thermus	15	21	6	13	3	48	26	97	96	5	24	1617	58	107	148	217	284	513	12	297	140	319	306	309	193	199	352	678
8.1. Hadobacteria	15	21	6	13	3	48	26	97	96	5	24	1617	58	107	148	217	284	513	12	297	140	319	306	309	193	199	352	678
8.1.1. Deinococcales	15	21	6	13	3	48	26	97	96	5	24	1617	58	107	148	217	284	513	12	297	140	319	306	309	193	199	352	678
8.1.1.1. Deinococcaceae	15	21	6	13	3	48	26	97	96	5	24	1617	58	107	148	217	284	513	12	297	140	319	306	309	193	199	352	678
8.1.1.1.1. Deinococcus	15	21	6	13	3	48	26	97	96	5	24	1617	58	107	148	217	284	513	12	297	140	319	306	309	193	199	352	678
8.1.2. Thermales	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	4	0	0	0	0	1	0	0	0	0	0	0	0
8.1.2.1. Thermaceae	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	4	0	0	0	0	1	0	0	0	0	0	0	0
8.1.2.1.1. Marinithermus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
8.1.2.1.2. Oceanithermus	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8.1.2.1.3. Thermus	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0
9. Firmicutes	155	31	23	29	100	122	1381	160	729	37	8	180	58	174	232	423	574	649	43	419	155	169	435	280	135	544	53	1380
9.1. Bacilli	155	31	23	29	100	122	1381	160	729	37	8	151	58	174	232	420	574	649	43	419	155	169	435	280	126	544	53	1380
9.1.1. Bacillales	155	31	23	29	100	122	1381	160	729	37	8	151	58	174	232	420	574	649	43	419	155	169	435	280	126	544	53	1380
9.1.1.1. Bacillaceae	155	31	23	27	100	122	1381	157	729	37	8	151	58	174	232	416	574	649	43	419	152	169	435	279	126	544	53	1380
9.1.1.1.1. Bacillus	153	31	23	27	100	122	1381	157	729	37	8	151	58	174	232	416	574	649	43	419	152	169	433	279	126	544	53	1380
9.1.1.1.2. Geobacillus	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
9.1.1.2. Paenibacillaceae	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	4	0	0	0	0	3	0	0	1	0	0	0	0
9.1.1.2.1. Thermobacillus	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	4	0	0	0	0	3	0	0	1	0	0	0	0
9.2. Clostridia	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
9.2.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
9.2.1.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
9.2.1.1.1. Symbiobacterium	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.1.2. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.1.2.1. Thermaerobacter	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.1.3. Peptococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
9.2.1.3.1. Desulfotomaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0
9.2.2. Thermoanaerobacterales	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.2.1. Thermoanaerobacteraceae	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.2.1.1. Thermacetogenium	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.2.1.2. Mahella	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.3. Negativicutes	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
9.3.1. Selenomonadales	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
9.3.1.1.. Acidominococcaceae	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
9.3.1.1..1. Selenomonas	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
10. Gemmatimonadetes	12	39	14	1	3	36	154	20	42	7	0	92	3	4	157	297	279	560	18	225	64	443	81	154	84	294	66	1552
10.1. Gemmatimonadetes	12	39	14	1	3	36	154	20	42	7	0	92	3	4	157	297	279	560	18	225	64	443	81	154	84	294	66	1552
10.1.1. Gemmantimonadales	12	39	14	1	3	36	154	20	42	7	0	92	3	4	157	297	279	560	18	225	64	443	81	154	84	294	66	1552
10.1.1.1. Gemmantimonadaceae	12	39	14	1	3	36	154	20	42	7	0	92	3	4	157	297	279	560	18	225	64	443	81	154	84	294	66	1552
10.1.1.1.1. Gemmatimonas	7	39	14	1	3	34	154	18	42	7	0	53	3	4	140	239	235	427	18	180	56	390	78	121	72	261	52	1395
10.1.1.1.2. Unclassified	5	0	0	0	0	2	0	2	0	0	0	39	0	0	17	58	44	133	0	45	8	53	3	33	12	33	14	157
11. Nitrospirae	0	0	0	0	0	0	6	0	2	0	6	0	0	0	0	13	2	3	12	0	0	0	0	2	0	0	0	0
11.1. Nitrospira	0	0	0	0	0	0	6	0	2	0	6	0	0	0	0	13	2	3	12	0	0	0	0	2	0	0	0	0
11.1.1. Nitrospirales	0	0	0	0	0	0	6	0	2	0	6	0	0	0	0	13	2	3	12	0	0	0	0	2	0	0	0	0
11.1.1.1. Nitrospiraceae	0	0	0	0	0	0	6	0	2	0	6	0	0	0	0	13	2	3	12	0	0	0	0	2	0	0	0	0
11.1.1.1.1. Nitrospira	0	0	0	0	0	0	6	0	2	0	6	0	0	0	0	13	2	3	12	0	0	0	0	2	0	0	0	0
12. Notyetclassified(Prokaryotae)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1. NC10phylumbacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1. CandidatusMethylomirabilis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1.1. CandidatusMethylomirabilis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1.1.1. Methylomirabilis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
13. Planctomycetes	214	126	33	103	133	151	203	426	113	105	2	317	17	28	574	1788	1025	862	136	636	720	578	859	951	1074	484	116	574
13.1. Phycisphaerae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	3	0	0	4	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
13.1.1. Phycisphaerales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	3	0	0	4	0	0	0	0
13.1.1.1. Phycisphaeraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	3	0	0	4	0	0	0	0
13.1.1.1.1. Phycisphaera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	3	0	0	4	0	0	0	0
13.2. Planctomycetacia	214	126	33	103	133	151	203	426	113	105	2	317	17	28	574	1788	1025	860	136	632	717	578	859	947	1074	484	116	574
13.2.1. Planctomycetales	214	126	33	103	133	151	203	426	113	105	2	317	17	28	574	1788	1025	860	136	632	717	578	859	947	1074	484	116	574
13.2.1.1. Planctomycetaceae	214	126	33	103	133	151	203	426	113	105	2	317	17	28	574	1788	1025	860	136	632	717	578	859	947	1074	484	116	574
13.2.1.1.1. Isosphaera	2	12	5	14	2	9	28	48	0	2	0	9	0	0	66	188	73	34	14	40	108	20	91	100	204	50	8	14
13.2.1.1.2. Pirellula	64	22	13	14	58	18	35	85	50	9	0	38	11	1	53	285	155	167	17	207	59	44	80	178	57	106	19	201
13.2.1.1.3. Planctomyces	0	0	0	0	0	0	0	0	9	0	0	0	3	0	3	0	0	15	0	20	17	0	0	3	23	2	7	12
13.2.1.1.4. Singulisphaera	148	92	15	75	73	124	140	293	54	94	2	270	3	27	452	1315	797	644	105	365	533	514	688	666	790	326	82	347
14. Proteobacteria	408	269	559	1644	3722	4644	349	2467	163	654	112	3406	96	88	1394	1952	1188	1598	1755	691	3332	1450	1426	1592	2659	593	104	1271
14.1. Alphaproteobacteria	90	121	194	414	1673	2354	94	1381	36	602	102	2248	38	43	368	788	307	332	924	183	1379	165	342	580	812	140	33	434
14.1.1. Caulobacteriales	22	18	0	10	13	44	0	333	5	274	94	1376	13	9	16	57	30	28	5	27	11	71	11	22	34	16	1	8
14.1.1.1. Caulobacteraceae	22	18	0	10	13	44	0	333	5	274	94	1376	13	9	16	57	30	28	5	27	11	71	11	22	34	16	1	8
14.1.1.1.1. Brevundimonas	0	0	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	10	0	1	0	0	0	0	0	3	1	0
14.1.1.1.2. Caulobacter	0	0	0	0	0	0	0	27	0	0	0	1	0	0	5	2	12	1	2	0	0	0	2	15	0	11	0	3
14.1.1.1.3. Phenylobacterium	22	18	0	10	13	44	0	304	5	274	94	1373	13	9	11	55	18	17	3	26	11	71	9	7	34	2	0	5
14.1.2. Hyphomicrobiales	65	92	160	399	1652	2294	94	1019	31	260	6	702	25	34	329	630	143	161	894	108	1352	68	310	261	726	81	8	110
14.1.2.1. Ancylobactergroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	0	0	1	0	0	0	0	0	0	0	0	0
14.1.2.1.1. Hyphomicrobium	0	0	0	0	0	0	0	0	0	0	0	3	0	0	14	1	0	0	1	0	0	0	0	0	0	0	0	0
14.1.2.2. Beijerinckiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	0	0	1	0	0	0	0	0	0	0	0	0
14.1.2.2.1. Methylocella	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
14.1.2.3. Bradyrhizobiaceae	33	67	80	194	824	1142	64	492	24	130	6	333	11	17	165	309	69	82	436	53	657	30	151	123	362	30	4	38
14.1.2.3.1. Bradyrhizobium	31	25	80	194	824	1142	29	488	7	128	0	332	11	17	126	297	65	77	430	42	655	30	147	123	298	30	4	34
14.1.2.3.2. Nitrobacter	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	19	0	0	0
14.1.2.3.3. Oligotropha	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	3	0	0	0	0	0
14.1.2.3.4. Rhodopseudomonas	0	42	0	0	0	0	35	4	17	2	6	1	0	0	39	9	1	5	4	11	2	0	1	0	45	0	0	4
14.1.2.4. Brucellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
14.1.2.4.1. Brucella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
14.1.2.5. Methylobacteriaceae	0	0	0	7	0	0	0	17	0	2	0	10	3	0	7	7	4	1	2	1	8	0	8	1	6	18	0	8
14.1.2.5.1. Methylobacterium	0	0	0	7	0	0	0	17	0	2	0	10	3	0	7	7	4	1	2	1	8	0	8	1	6	18	0	8
14.1.2.6. Methylocystaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	2
14.1.2.6.1. Methylocystis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	2

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
14.1.2.7. Phyllobacteriaceae	0	0	0	0	0	2	0	4	0	0	0	7	0	0	0	3	0	0	11	0	8	0	0	0	13	0	0	0
14.1.2.7.1. Mesorhizobium	0	0	0	0	0	2	0	4	0	0	0	7	0	0	0	3	0	0	11	0	8	0	0	0	13	0	0	0
14.1.2.8. Rhizobiaceae	31	25	80	198	828	1150	29	506	7	128	0	348	11	17	126	303	69	77	441	54	670	38	151	135	315	32	4	62
14.1.2.8.1. Agrobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	4	0	0	0	0	2
14.1.2.8.2. Neorhizobium	0	0	0	2	2	3	0	1	0	0	0	0	0	0	0	0	1	0	0	0	2	4	0	6	0	0	0	0
14.1.2.8.3. Rhizobium	31	25	80	196	826	1147	29	499	7	128	0	345	11	17	126	302	66	77	441	48	668	34	147	129	315	32	4	47
14.1.2.8.4. Sinorhizobium	0	0	0	0	0	0	0	6	0	0	0	3	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	13
14.1.2.9. Rhodobiaceae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0
14.1.2.9.1. Parvibaculum	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
14.1.2.9.2. Methyloceanibacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
14.1.2.10. Xanthobacteraceae	0	0	0	0	0	0	1	0	0	0	0	4	0	0	3	6	1	0	2	0	9	0	0	0	0	0	0	0
14.1.2.10.1. Azorhizobium	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.1.2.10.2. Starkeya	0	0	0	0	0	0	1	0	0	0	0	1	0	0	3	3	1	0	2	0	8	0	0	0	0	0	0	0
14.1.2.10.3. Xanthobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0
14.1.3. Rhodobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	32	0	0	0
14.1.3.1. Rhodobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	32	0	0	0
14.1.3.1.1. Roseibacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	32	0	0	0
14.1.4. Rhodospirillales	0	0	3	0	3	0	0	4	0	0	0	4	0	0	3	90	114	22	2	10	1	22	19	291	0	32	4	33
14.1.4.1. Acetobacteraceae	0	0	3	0	3	0	0	4	0	0	0	4	0	0	3	90	114	22	2	10	1	22	19	291	0	32	4	33
14.1.4.1.1. Acidiphilium	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.1.4.1.2. Gluconacetobacter	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.1.4.1.3. Gluconobacter	0	0	3	0	3	0	0	0	0	0	0	2	0	0	3	90	114	22	2	10	1	22	19	291	0	32	4	33
14.1.5. Rhodospirillales	1	2	0	0	0	0	0	1	0	0	0	3	0	0	10	7	3	4	1	0	3	1	0	1	14	4	0	5
14.1.5.1. Rhodospirillaceae	1	2	0	0	0	0	0	1	0	0	0	3	0	0	10	7	3	4	1	0	3	1	0	1	14	4	0	5
14.1.5.1.1. Azospirillum	0	0	0	0	0	0	0	1	0	0	0	1	0	0	6	4	1	1	1	0	0	0	0	0	10	0	0	1
14.1.5.1.2. Magnetospirillum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	0	0	0
14.1.5.1.3. Rhodocista	1	2	0	0	0	0	0	0	0	0	0	2	0	0	4	3	2	0	0	0	2	1	0	1	3	4	0	4
14.1.5.1.4. Rhodospirillum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.1.6. Sphingomonadales	2	9	31	5	5	16	0	24	0	68	2	163	0	0	10	4	17	117	22	38	8	3	2	5	6	7	20	278
14.1.6.1. Sphingomonadaceae	2	9	31	5	5	16	0	24	0	68	2	163	0	0	10	4	17	117	22	38	8	3	2	5	6	7	20	278
14.1.1.1. Sphingobium	0	1	0	0	0	0	0	10	0	1	2	3	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0
14.1.1.1.2. Sphingomonas	2	2	31	4	5	16	0	7	0	67	0	160	0	0	8	4	17	116	22	36	8	3	0	3	6	7	20	276
14.1.1.1.3. Sphingopyxis	0	6	0	1	0	0	0	7	0	0	0	0	0	0	0	0	0	1	0	2	0	0	2	0	0	0	0	2
14.2. Betaproteobacteria	47	12	17	44	54	75	35	70	10	6	1	124	18	15	168	223	159	188	135	49	188	88	506	69	123	70	26	267

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
14.2.1. Burkholderiales	44	12	17	44	54	69	35	70	10	6	0	121	18	15	163	220	153	184	130	49	156	88	506	65	105	70	26	266
14.2.1.1. Alcaligenaceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	32	4	0	0	0	3	1	0	1	0	0	0	0
14.2.1.1.1. Achromobacter	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
14.2.1.1.2. Alcaligenes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	4	0	0	0	1	0	0	0	0	0	0	0
14.2.1.1.3. Bordetella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	1	0	0	0	0
14.2.1.2. Burkholderiaceae	0	0	0	0	0	0	15	3	0	0	0	42	5	0	84	31	36	14	59	11	53	37	38	7	58	32	0	67
14.2.1.2.1. Acinetobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
14.2.1.2.2. Burkholderia	0	0	0	0	0	0	0	1	0	0	0	24	0	0	4	19	10	0	15	0	49	0	10	6	56	1	0	3
14.2.1.2.3. Cupriavidus	0	0	0	0	0	0	8	0	0	0	0	6	0	0	5	0	10	0	1	0	0	0	0	0	2	2	0	5
14.2.1.2.4. Pandoraea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
14.2.1.2.5. Ralstonia	0	0	0	0	0	0	7	2	0	0	0	12	5	0	75	12	16	14	43	11	3	37	28	1	0	29	0	51
14.2.1.3. Comamonadaceae	44	10	15	44	44	66	1	67	0	5	0	74	11	15	63	151	102	115	18	30	62	29	379	46	17	28	17	45
14.2.1.3.1. Acidivorax	10	7	0	12	4	3	0	7	0	0	0	59	8	15	4	45	67	55	4	9	17	11	17	13	9	4	0	2
14.2.1.3.2. Acidovorax	16	0	0	2	0	0	1	0	0	3	0	0	0	0	0	1	0	1	0	0	0	0	0	6	1	0	0	2
14.2.1.3.3. Albidiferax	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	32	8	25	1	2	2	12	9	13	2	13	0	33
14.2.1.3.4. Alicycliphilus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
14.2.1.3.5. Comamonas	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.2.1.3.6. Curvibacter	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
14.2.1.3.7. Delftia	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	13	0	0	0	0	1	0	1
14.2.1.3.8. Polaromonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
14.2.1.3.9. Variovorax	3	3	15	30	40	37	0	60	0	2	0	14	0	0	48	71	27	32	13	19	30	4	11	14	3	10	17	6
14.2.1.3.10. Verminephrobacter	5	0	0	0	0	26	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	342	0	0	0	0	1
14.2.1.4. Oxalobacteraceae	0	2	0	0	10	3	19	0	6	0	0	3	2	0	10	0	8	53	48	6	28	20	88	11	30	4	7	153
14.2.1.4.1. Collimonas	0	2	0	0	10	3	19	0	0	0	0	0	2	0	10	0	8	53	48	6	27	16	62	11	22	4	7	153
14.2.1.4.2. Herbaspirillum	0	0	0	0	0	0	0	0	6	0	0	3	0	0	0	0	0	0	0	0	0	4	26	0	8	0	0	0
14.2.1.4.3. Janthinobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.2.1.5. Unclassified Burkholderiales	0	0	2	0	0	0	0	0	4	1	0	0	0	0	6	6	3	2	5	2	10	1	1	0	0	6	2	1
14.2.1.5.1. Leptothrix	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	1	10	0	0	0	0	0	0	1
14.2.1.5.2. Methylibium	0	0	0	0	0	0	0	0	1	1	0	0	0	0	6	6	1	2	3	1	0	1	1	0	0	6	2	0
14.2.1.5.3. Rhodocystis	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.2.2. Gallionellales	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
14.2.2.1. Gallionellagroup	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
14.2.2.1.1. Gallionella	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.2.2.1.2. Sideroxydans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
14.2.3. Hydrogenophilales	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	2	0	2	0	0	2	0	0	1	0	0	0	0
14.2.3.1. Hydrogenophilaceae	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	2	0	2	0	0	2	0	0	1	0	0	0	0
14.2.3.1.1. Thiobacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	1	0	0	0	0
14.2.4. Methylophilales	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0
14.2.4.1. Methylophilaceae	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0
14.2.4.1.1. Methylotenera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
14.2.5. Neisseriales	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
14.2.5.1. ChromobacteriaceaeA	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
14.2.5.1.1. Chromobacterium	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.2.4.1.2. Pseudogulbenkiania	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
14.2.5.2. Neisseriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.2.5.2.1. Neisseria	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.2.6. Nitrosomonadales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.2.6.1. Nitrosomonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.2.6.1.1. Nitrosospira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
14.2.7. Rhodocyclales	3	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	2	1	0	0	26	0	0	2	18	0	0	1
14.2.7.1. Rhodocyclaceae	3	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	2	1	0	0	26	0	0	2	18	0	0	1
14.2.7.1.1. Aromatoleum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1
14.2.7.1.2. Azoarcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	0	0	0	0
14.2.7.1.3. Azospira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
14.2.7.1.4. Sulfuritalea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
14.2.7.1.5. Thauera	3	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	1	0	0	14	0	0	0	18	0	0	0
14.2.8. Sulfuricellales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.2.8.1. Sulfuricellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.2.8.1.1. Sulfuricella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0
14.2.9. Unclassified Betaproteobacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	3	0	0	0	0	0	0	0	0	0
14.2.9.1. Candidatus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	3	0	0	0	0	0	0	0	0	0
14.2.9.1.1. Accumulibacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	3	0	0	0	0	0	0	0	0	0
14.3. Gammaproteobacteria	271	133	348	1186	1995	2215	220	1016	117	46	9	1034	40	30	852	920	711	1072	690	459	1759	1197	569	943	1724	383	45	570
14.3.1. Acidithiobacillales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	0	0	0	0
14.3.1.1. Acidithiobacillaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	1	0	0	0	0
14.3.1.1.1. Acidithiobacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0
14.3.1.1.2. Marinobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
14.3.2. Chromatiaceae/Ectothiorhodospiraceae group	0	0	0	0	0	0	0	0	0	7	0	2	0	0	0	0	2	13	1	0	0	0	2	0	0	0	0	11

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
14.3.2.1. Chromatiaceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	13	0	0	0	0	2	0	0	0	0	11
14.3.2.1.1. Allochromatium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	2	0	0	0	0	11
14.3.2.1.2. Candidatus	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.3.2.1.3. Marichromatium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
14.3.2.1.4. Ectothiorhodospira group	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.3.2.1.5. Halorhodospira	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.3.2.1.6. Ectothiorhodospira group	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.3.2.1.7. Thioalkalivibrio	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
14.3.3. Methylococcaceae group	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
14.3.3.1. Methylococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
14.3.3.1.1. Methylococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
14.3.4. Oceanospirillales	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	5
14.3.4.1. Alcanivorax/Fundibacter group	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	5
14.3.4.1.1. Chromohalobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
14.3.4.1.2. Halomonas	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	2
14.3.5. Pseudomonaceae/Moraxellaceae group	122	99	222	1086	1342	1873	128	446	62	30	6	487	13	10	457	692	591	904	214	343	957	1000	552	822	1061	271	44	479
14.3.5.1. Pseudomonadaceae	122	99	222	1086	1342	1873	128	446	62	30	6	487	13	10	457	692	591	904	214	343	957	1000	552	822	1061	271	44	479
14.3.5.1.1. Azotobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	2	0	2	2	0	0	0
14.3.5.1.2. Chlorobacterium	69	33	71	650	1004	1377	6	302	18	11	0	346	11	7	211	436	255	698	171	180	523	822	514	679	700	147	34	194
14.3.5.1.3. Pseudomonas	53	66	151	436	338	496	122	144	44	19	6	141	2	3	242	256	336	206	43	163	434	176	38	141	359	124	10	285
14.3.6. Xanthomonadales	149	34	126	100	653	342	92	570	53	9	3	545	27	20	395	226	118	154	475	116	802	197	13	120	661	112	1	75
14.3.6.1. Lysobacteraceae	149	34	126	100	653	342	92	570	53	9	3	545	27	20	395	226	118	154	475	116	802	197	13	120	661	112	1	75
14.3.6.1.1. Dyella	3	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
14.3.6.1.2. Lysobacter	0	0	0	0	0	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14.3.6.1.3. Pseudoxanthomonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0
14.3.6.1.4. Rhodanobacter	4	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
14.3.6.1.5. Stenotrophomonas	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	6	7	0	1	1	0	0	0	0	5	1	1
14.3.6.1.6. Xanthomonas	142	34	126	96	653	342	0	570	53	9	3	542	27	20	395	226	112	147	475	115	790	197	13	120	661	107	0	74
14.4. subdelta/epsilonsubdivisions	0	3	0	0	0	0	0	0	0	0	0	0	0	0	6	21	11	6	6	0	6	0	9	0	0	0	0	0
14.4.1. Campylobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.1.1. Unclassified Campylobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.1.1.1. Nitratifractor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
14.4.2. Desulfobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	1	0	6	0	6	0	9	0	0	0	0	0
14.4.2.1. Desulfobulbaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	1	0	6	0	6	0	9	0	0	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
14.4.2.1.1. Desulfurivibrio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6	0	5	0	0	0	3	0	0	0
14.4.2-1-2. Desulfovibrio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	0	0	0	0	0	0	0	0	0	0	0
14.4.3. Desulfuromonales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.3.1. Geobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.3.1.1. Geoalkalibacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.3.2. Geobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
14.4.3.2.1. Geobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0
14.4.4. Myxobacteria	0	3	0	0	0	1	2	3	0	0	0	33	0	0	4	13	8	6	0	0	8	2	9	9	14	0	0	11
14.4.4.1. Anaeromyxobacteraceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	0	2	0	0	2	2	0	0	2	0	0	2
14.4.4.1.1. Anaeromyxobacter	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	0	2	0	0	2	2	0	0	2	0	0	2
14.4.4.2. Archangiaceae	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0
14.4.4.2.1. Stigmatella	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0
14.4.4.3. Haliangiaceae	0	0	0	0	0	1	2	0	0	0	0	3	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
14.4.4.3.1. Haliangium	0	0	0	0	0	1	2	0	0	0	0	3	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
14.4.4.4. Myxococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
14.4.4.4.1. Corallococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
14.4.4.5. Myxococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0	6	0	9	0	0	0	0	0
14.4.4.5.1. Myxococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0	6	0	9	0	0	0	0	0
14.4.4.6. Sorangiaceae	0	3	0	0	0	0	0	3	0	0	0	27	0	0	4	3	7	0	0	0	0	0	0	9	0	0	0	9
14.4.4.6.1. Sorangium	0	3	0	0	0	0	0	3	0	0	0	27	0	0	4	3	7	0	0	0	0	0	0	9	0	0	0	9
14.4.5. Syntrophobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.5.1. Syntrophaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.5.1.1. Desulfobacca	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.6. Syntrophobacterales	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.6.1. Syntrophobacteraceae	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.6.1.1. Syntrophobacter	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14.4.7. Spirochaetales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	2	0	0	0	0
14.4.7.1. Spirochaetaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	2	0	0	0	0
14.4.7.1.1. Spirochaeta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	2	0	0	0	0	0	0	2	0	0	0	0
15. Eubacteria environmental samples	67	2	138	52	128	186	52	49	16	35	1	115	21	36	156	103	253	96	59	187	201	183	99	123	263	82	26	549
Fungi	0	0	0	0	0	12	0	4	0	0	0	4	0	0	1	0	0	0	2	0	34	0	45	0	39	0	0	0
16. Ascomycota	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	2	0	34	0	45	0	39	0	0	0

Table S4.4. Taxonomic summary of the *phoD* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Guf	AUS2-F	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-F	CH5-G
16.1. Arthrodermataceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
16.1.1. Arthroderma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
16.2. Botryosphaeriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
16.2.1. Neofusicoccum	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.3. Capnodiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
16.3.1. Baudoinia	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.3.2. Mycosphaerella	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4. Chaetomiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
16.4.1. Chrysosporium	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.5. Eurotiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
16.5.1. Penicillium	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.6. Herpotrichiellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	9	0	0	0	0	0
16.6.1. Capronia	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.7. Hypocreaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0
16.7.1. Trichoderma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	0	0
16.8. Hypocreales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
16.8.1. Hypocrea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
16.9. Melanommatales	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.9.1. Leptosphaeria	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17. Basidiomycota	0	0	0	0	0	12	0	4	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17.1. Mycosyringaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17.1.1. Pseudozyma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
17.2. Serpulaceae	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.2.1. Serpula	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.3. Stereaceae	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.3.1. Stereum	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.4. Tremellaceae	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17.4.1. Cryptococcus	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples.

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
Total	3470	9318	177	968	389	1911	863	2513	2093	2796	568	1626	4694	1154	2335	7839	3249	7114	109	2236	1071	7020	641	4504	34	3335	10489
Archaea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
1. Euryarchaeota	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
1.1. Methanomicrobia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
1.1.1. Halomebacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
1.1.1.1. Methanothrix	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
1.1.1.1.1. Methanosaeta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0
Bacteria	3470	9318	177	968	389	1911	863	2513	2093	2796	568	1626	4694	1154	2335	7839	3241	7114	109	2236	1071	7020	641	4504	34	3335	10489
2. Acidobacteria	17	14	99	21	0	0	134	4	0	2	7	0	0	0	4	9	1	10	19	23	0	2	11	0	0	33	3
2.1. Acidobacteria	17	14	99	21	0	0	134	4	0	2	7	0	0	0	4	9	1	10	19	23	0	2	11	0	0	33	3
2.1.1. Acidobacteria	17	14	99	21	0	0	134	4	0	2	7	0	0	0	4	9	1	10	19	23	0	2	11	0	0	33	3
2.1.1.1. Acidobacteriaceae	16	10	99	1	0	0	124	0	0	2	2	0	0	0	0	7	1	10	19	1	0	2	9	0	0	21	3
2.1.1.1.1. Acidobacterium	0	7	99	1	0	0	124	0	0	2	2	0	0	0	0	4	0	10	19	1	0	2	9	0	0	21	1
2.1.1.1.2. Terriglobus	16	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	0	0	0	2
2.1.1.2. Candidatus	1	4	0	20	0	0	10	4	0	0	5	0	0	0	4	2	0	0	0	22	0	0	2	0	0	12	0
2.1.1.2.1. Chloracidobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
2.1.1.2.2. Unclassified acidobacterium	1	4	0	20	0	0	10	4	0	0	5	0	0	0	4	2	0	0	0	22	0	0	2	0	0	5	0
3. Actinobacteria	1618	3957	59	0	9	596	349	685	549	1568	384	411	1042	134	212	591	1278	1836	15	397	111	1276	5	879	5	489	2312
3.1. Acidimicrobiales	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1.1. Acidimicrobiaceae	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1.1.1. Ilumatobacter	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2. Actinomycetales	1613	872	49	0	9	594	347	559	313	1493	332	396	1005	132	198	567	1175	1821	12	397	111	1245	3	820	5	487	2163
3.2.1. Actinoplanaceae	60	77	0	0	0	19	0	33	0	34	0	54	123	2	26	57	96	64	0	0	0	29	0	114	5	54	238
3.2.1.1. Actinomyces	4	0	0	0	0	0	0	3	0	0	0	1	0	0	23	31	8	1	0	0	0	11	0	4	0	8	0
3.2.1.2. Actinoplanes	0	52	0	0	0	1	0	8	0	3	0	0	27	2	0	3	23	8	0	0	0	0	0	35	5	1	38
3.2.1.3. Ampullaria	2	0	0	0	0	0	0	4	0	0	0	1	15	0	0	0	30	0	0	0	0	0	0	0	0	1	29
3.2.1.4. Kitasatospora	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.1.5. Micromonospora	2	0	0	0	0	7	0	1	0	4	0	0	18	0	0	0	5	3	0	0	0	0	0	4	0	0	29
3.2.1.6. Microstreptospora	22	25	0	0	0	7	0	4	0	7	0	40	32	0	0	17	4	25	0	0	0	7	0	5	0	12	81
3.2.1.7. Salinospora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
3.2.1.8. Streptomyces	20	0	0	0	0	4	0	4	0	17	0	12	31	0	3	0	12	0	0	0	0	10	0	40	0	22	28
3.2.1.9. Streptosporangium	0	0	0	0	0	0	0	5	0	3	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	3	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
3.2.1.10. Streptotrix	9	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	15
3.2.1.11. Thermomonospora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	26	0	7	9
3.2.1.12. Verrucosispora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	27	0	0	0	1	0	0	0	0	8
3.2.2. Actinosynnemataceae	8	0	0	0	0	43	25	111	2	127	34	17	99	5	0	31	243	60	0	1	0	10	0	24	0	21	263
3.2.2.1. Actinobispora	0	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.2.2. Actinosynnema	2	0	0	0	0	0	0	14	0	64	0	2	0	0	0	6	63	49	0	0	0	1	0	19	0	0	23
3.2.2.3. Amycolatopsis	2	0	0	0	0	0	25	3	0	4	0	0	0	0	0	0	9	11	0	0	0	1	0	0	0	0	16
3.2.2.4. Faenia	0	0	0	0	0	17	0	3	0	5	0	13	0	0	0	0	39	0	0	1	0	1	0	1	0	0	0
3.2.2.5. Kutzneria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
3.2.2.6. Saccharomonospora	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.2.7. Saccharothrix	4	0	0	0	0	26	0	91	2	54	0	0	99	5	0	24	132	0	0	0	0	7	0	4	0	19	224
3.2.2.8. Thermopolyspora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
3.2.3. Beutenbergiaceae	2	0	0	0	0	0	0	4	0	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0
3.2.3.2. Beutenbergia	2	0	0	0	0	0	0	4	0	0	0	0	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0
3.2.4. Cellulomonadaceae	20	0	0	0	0	2	0	0	0	8	0	0	6	13	0	5	112	0	0	0	0	0	0	10	0	0	30
3.2.4.1. Aplanobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
3.2.4.2. Cellulomonas	20	0	0	0	0	2	0	0	0	8	0	0	6	13	0	5	107	0	0	0	0	0	0	10	0	0	30
3.2.5. Corynebacteriaceae	1	0	0	0	0	1	1	0	0	1	2	0	1	0	0	5	14	1	0	0	0	2	0	0	0	0	10
3.2.5.1. Caseobacter	0	0	0	0	0	0	1	0	0	1	2	0	0	0	0	0	14	0	0	0	0	2	0	0	0	0	6
3.2.5.2. Corynebacterium	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	4
3.2.5.3. Flavobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0
3.2.6. Dermabacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3.2.6.1. Brachybacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
3.2.7. Dermacoccaceae	2	9	0	0	0	0	0	0	0	0	2	0	0	1	0	0	2	0	0	0	0	0	0	6	0	0	0
3.2.7.1. Dermacoccus	2	9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0
3.2.7.2. Kytococcus	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
3.2.8. Frankiaceae	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
3.2.8.1. Frankia	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11
3.2.9. Geodermatophilaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3.2.9.1. Geodermatophilus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3.2.10. Gordoniaceae	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.10.1. Gordona	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.11. Intrasporangiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
3.2.11.1. Humihabitans	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0
3.2.12. Kineosporiaceae	0	3	0	0	0	0	0	0	0	0	0	0	3	0	0	2	6	0	0	0	0	0	0	1	0	57	4
3.2.12.1. Kineococcus	0	3	0	0	0	0	0	0	0	0	0	0	3	0	0	2	6	0	0	0	0	0	0	1	0	57	4

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
3.2.13. Microbacteriaceae	9	0	0	0	0	15	0	0	2	1	0	0	9	8	6	35	60	0	0	0	0	28	0	8	0	4	63
3.2.13.2. Clavibacter	9	0	0	0	0	15	0	0	2	1	0	0	9	8	6	35	60	0	0	0	0	28	0	8	0	4	63
3.2.14. Micrococcaceae	1250	0	27	0	0	346	1	36	216	125	38	37	593	0	3	33	186	0	7	0	3	78	0	40	0	15	76
3.2.14.1. Arthrobacter	1224	0	27	0	0	331	1	34	216	119	38	37	586	0	3	33	49	0	2	0	3	78	0	33	0	15	58
3.2.14.2. Kocuria	0	0	0	0	0	15	0	1	0	1	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	13
3.2.14.3. Micrococcus	26	0	0	0	0	0	0	1	0	5	0	0	7	0	0	0	131	0	5	0	0	0	0	7	0	0	5
3.2.15. Microsphaeraceae	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	8	16	0	0	0	0	13	0	2	0	0	79
3.2.15.1. Humicoccus	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	8	16	0	0	0	0	13	0	2	0	0	79
3.2.16. Mycobacteriaceae	224	437	22	0	9	128	320	367	88	1098	255	173	84	68	106	286	254	1600	5	344	37	785	3	436	0	226	1250
3.2.16.1. Mycobacterium	117	437	22	0	9	36	320	345	83	1041	255	113	5	68	105	265	210	1524	5	344	37	756	3	378	0	158	860
3.2.16.2. Nocardia	107	0	0	0	0	92	0	22	5	57	0	60	79	0	1	21	44	76	0	0	0	29	0	58	0	68	390
3.2.17. Nocardiaceae	18	208	0	0	0	37	0	1	2	1	0	0	0	0	0	7	9	81	0	0	0	5	0	48	0	37	24
3.2.17.1. Micropolyspora	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
3.2.17.2. Rhodococcus	18	208	0	0	0	37	0	1	2	1	0	0	0	0	0	4	9	81	0	0	0	5	0	48	0	37	24
3.2.18. Nocardioidaceae	0	1	0	0	0	2	0	4	0	10	0	9	33	0	0	7	18	9	0	1	0	0	0	2	0	3	12
3.2.18.1. Hongia	0	0	0	0	0	2	0	4	0	4	0	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.18.2. Nocardioides	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	11	0	0	0	0	0	0	0	0	0	3
3.2.18.3. Brachystreptospora	0	1	0	0	0	0	0	0	0	0	0	5	3	0	0	2	4	1	0	1	0	0	0	2	0	3	9
3.2.18.4. Thermonospora	0	0	0	0	0	0	0	0	0	6	0	2	26	0	0	5	3	8	0	0	0	0	0	0	0	0	0
3.2.19. Promicromonosporaceae	9	56	0	0	0	1	0	0	0	4	0	0	0	0	0	0	28	0	0	0	0	37	0	0	0	0	4
3.2.19.1. Isoptericola	9	56	0	0	0	1	0	0	0	2	0	0	0	0	0	0	28	0	0	0	0	37	0	0	0	0	4
3.2.19.2. Xylanimonas	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2.20. Propionibacteriaceae	8	81	0	0	0	0	0	3	0	84	0	106	53	33	56	91	128	6	0	51	71	258	0	116	0	70	99
3.2.20.1. Microlunatus	8	81	0	0	0	0	0	3	0	84	0	106	53	33	56	91	128	6	0	51	71	258	0	116	0	70	99
3.2.21. Tsukamurellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
3.2.21.1. Tsukamurella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0
3.3. Bifidobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0
3.3.1. Bifidibacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0
3.3.1.1. Bifidobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	0	0	0	0
3.4. Rubrobacterales	3	0	0	0	0	0	2	13	16	60	50	15	34	0	0	0	2	0	0	0	0	5	0	14	0	0	80
3.4.1. Rubrobacteraceae	3	0	0	0	0	0	2	13	16	60	50	15	34	0	0	0	2	0	0	0	0	5	0	14	0	0	80
3.4.1.1. Rubrobacter	3	0	0	0	0	0	2	13	16	60	50	15	34	0	0	0	2	0	0	0	0	5	0	14	0	0	80
3.5. Solirubrobacterales	2	3085	2	0	0	2	0	113	220	15	2	0	3	2	14	24	101	15	0	0	0	26	0	45	0	2	69
3.5.1. Conexibacteraceae	2	3085	2	0	0	2	0	113	220	15	2	0	3	2	14	24	101	15	0	0	0	26	0	45	0	2	69
3.5.1.1. Conexibacter	2	9	0	0	0	2	0	113	220	15	2	0	3	2	1	24	101	0	0	0	0	26	0	45	0	2	69
3.5.1.2. unclassiedSolibacteraceae	0	3076	2	0	0	0	0	0	0	0	0	0	0	0	13	0	0	15	0	0	0	0	0	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
4. Aminanaerobia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	3	0
4.1. Synergistia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	3	0
4.1.1. Synergistales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	3	0
4.1.1.1. Anaerobaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	3	0
4.1.1.1.1. Anaerobaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	0	0	0	3	0
5. Armatimonadetes	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	14
5.1. Chthonomonadetes	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5.1.1. Chthonomonadales	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5.1.1.1. Chthonomonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5.1.1.1.1. Chthonomonas	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5.2. Fimbriimonadia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	14
5.2.1. Fimbriionadales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	14
5.2.1.1. Fimbriionadales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	14
5.2.1.1.1. Fimbriimonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	14
6. Bacteroidetes	0	12	1	0	0	0	4	5	4	11	1	0	0	0	1	2	0	2	0	1	2	0	0	9	0	3	8
6.1. Bacteroidia	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	2	0	0	0	0	0	0
6.1.1. Bacteroidales	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	2	0	0	0	0	0	0
6.1.1.1. Rhodothermaceae	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
6.1.1.1.1. Rhodothermus	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
6.1.1.2. Prolixibacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0
6.1.1.2.1. Draconibacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0
6.2. Cytophagia	0	9	1	0	0	0	0	1	0	7	0	0	0	0	0	2	0	0	0	1	0	0	0	9	0	3	2
6.2.1. Cytophagales	0	9	1	0	0	0	0	1	0	7	0	0	0	0	0	2	0	0	0	1	0	0	0	9	0	3	2
6.2.1.1. Cytophagaceae	0	9	1	0	0	0	0	1	0	7	0	0	0	0	0	2	0	0	0	0	0	0	0	9	0	3	2
6.2.1.1.1. Dyadobacter	0	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1
6.2.1.1.2. Fibrella	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.2.1.1.3. Runella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3	1
6.2.1.1.4. Spirosoma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
6.2.1.1.5. Taxeobacter	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6.2.1.2. Flammeovirgaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6.2.1.2.1. Marivirga	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6.3. Sphingobacteriia	0	0	0	0	0	0	4	4	4	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
6.3.1. Sphingobacteriales	0	0	0	0	0	0	4	4	4	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
6.3.1.1. Chitinophagaceae	0	0	0	0	0	0	2	4	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
6.3.1.1.1. Chitiniphaga	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
6.3.1.1.2. Niabella	0	0	0	0	0	0	0	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3.1.1.3. Niastella	0	0	0	0	0	0	2	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3.1.2. Saprospiraceae	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3.1.2.1. Haliscomenobacter	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3.1.3. Sphingobacteriaceae	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6.3.1.3.1. Pedobacter	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7. CandidatusSaccharibacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7.1. CandidatusSaccharibacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
8. Chlorobi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
8.1. Chlorobia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
8.1.1. Chlorobiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
8.1.1.1. Chlorobiacea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
8.1.1.1.1. Chlorobaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0
9. Firmicutes	0	0	4	0	0	0	11	0	2	0	0	0	0	0	0	0	3	11	0	0	0	0	0	0	0	0	0
9.1. Bacilli	0	0	4	0	0	0	11	0	2	0	0	0	0	0	0	0	3	11	0	0	0	0	0	0	0	0	0
9.1.1. Bacillales	0	0	4	0	0	0	11	0	2	0	0	0	0	0	0	0	3	11	0	0	0	0	0	0	0	0	0
9.1.1.1. Bacillaceae	0	0	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
9.1.1.1.1. Bacillus	0	0	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0
9.1.1.2. Paenibacillaceae	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
9.1.1.2.1. Geobacillus	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.1.1.2.2. Thermobacillus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
9.2. Clostridia	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	2	2	0	0	1	0	0	0	0	0	0
9.2.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0
9.2.1.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
9.2.1.1.1. Aerothermobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
9.2.1.2. Ruminococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
9.2.1.2.1. Ethanoligenens	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
9.2.2. Halanaerobiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
9.2.2.1. Halanaerobiaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
9.2.2.1.1. Halothermothrix	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
9.2.3. Thermoanaerobacter group	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.3.1. Thermoanaerobacteraceae	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9.2.3.1.1. Clostridium	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
9.3. Unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
10. Chloroflexi	6	122	0	0	0	0	0	6	1385	36	50	27	10	0	0	12	2	5	0	0	0	22	0	2	0	2	19
10.1. Anaerolineae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
10.1.1. Anaerolinaeles	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
10.1.1.1. Anaerolinaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
10.1.1.1.1. Anaerolinea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
10.2. Chloroflexi	2	122	0	0	0	0	0	6	1385	34	50	27	10	0	0	12	2	3	0	0	0	22	0	2	0	2	19
10.2.1. Chloroflexaceae	2	122	0	0	0	0	0	6	1153	16	0	2	10	0	0	12	2	3	0	0	0	22	0	2	0	2	19
10.2.1.1. Chloroflexaceae	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.2.1.1.1. Chlorocrinis	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.2.1.1.2. Chloroflexus	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.2.1.2. Roseiflexaceae	2	122	0	0	0	0	0	4	1153	16	0	1	10	0	0	12	2	3	0	0	0	22	0	2	0	2	19
10.2.1.2.1. Roseiflexus	2	122	0	0	0	0	0	4	1153	16	0	1	10	0	0	12	2	3	0	0	0	22	0	2	0	2	19
10.2.2. Herpetosiphonales	0	0	0	0	0	0	0	0	232	18	50	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.2.2.1. Herpetosiphonaceae	0	0	0	0	0	0	0	0	232	18	50	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.2.2.1.1. Herpetosiphon	0	0	0	0	0	0	0	0	232	18	50	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3. Thermomicrobia	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.1. Sphaerobacterales	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.1.1. Sphaerobacteraceae	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.1.1.1. Sphaerobacter	4	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.2. Thermomicrobiales	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.2.1. Thermomicrobiaceae	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10.3.2.1.1. Thermomicrobium	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11. Cyanobacteria	0	0	0	0	0	0	0	10	0	1	0	0	40	0	0	0	11	0	0	0	0	0	0	9	0	4	32
11.1. CyanobacteriaSubsectionII	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.1.1. Pleurocapsales	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.1.1.1. Chroococcidiopsis	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.2. Gloeobacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	9	0	4	32
11.2. Gloeobacteria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	9	0	4	32
11.2.1. Gloeobacterales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	9	0	4	32
11.2.1.1. Gloeobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	9	0	4	32
11.2.2. Synechococcales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
11.2.2.1. Synechococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
11.3. Oscillatoriophycideae	0	0	0	0	0	0	0	10	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
11.3.1. Oscillatoriales	0	0	0	0	0	0	0	10	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.3.1.1. Oscillatoriaceae	0	0	0	0	0	0	0	10	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.3.1.1.1. Geitlerinema	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11.3.1.1.2. Oscillatoria	0	0	0	0	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12. Deinococcus-Thermus	6	9	0	4	0	0	2	0	1	3	0	10	8	0	0	2	4	0	0	0	0	0	0	1	0	5	8
12.1. Hadobacteria	6	9	0	4	0	0	2	0	1	3	0	10	8	0	0	2	4	0	0	0	0	0	0	1	0	5	8
12.1.1. Deinococcales	6	0	0	4	0	0	2	0	1	3	0	10	8	0	0	0	2	0	0	0	0	0	0	1	0	2	8
12.1.1.1. Deinococcaceae	6	0	0	4	0	0	2	0	1	3	0	8	8	0	0	0	2	0	0	0	0	0	0	1	0	2	8
12.1.1.1.1. Deinococcus	6	0	0	4	0	0	2	0	1	3	0	8	8	0	0	0	2	0	0	0	0	0	0	1	0	2	8
12.1.1.2. Trueperaceae	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1.1.2.1. Truepera	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12.1.2. Thermales	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	3	0
12.1.2.1. Thermaceae	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	3	0
12.1.2.1.1. Meiothermus	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
12.1.2.1.2. Oceanithermus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
12.1.2.1.3. Thermus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
13. Gemmatimonadetes	2	26	0	0	0	2	0	4	0	11	2	0	0	0	0	9	0	2	0	0	0	0	0	4	0	0	0
13.1. Unclassified	2	26	0	0	0	2	0	4	0	11	2	0	0	0	0	9	0	2	0	0	0	0	0	4	0	0	0
14. Nitrospirae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0
14.1. Nitrospira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0
14.1.1. Nitrospirales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0
14.1.1.1. Nitrospiraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0
14.1.1.1.1. Nitrospira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0
15. Planctomycetes	3	2	0	0	0	1	6	3	3	14	25	138	41	9	3	69	54	5	2	0	1	4	0	76	0	365	404
15.1. Phycisphaerae	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.1.1. Phycisphaerales	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.1.1.1. Phycisphaeraceae	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15.1.1.1.1. Phycisphaera	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16. Planctomycetacia	0	2	0	0	0	1	6	3	3	14	25	138	41	9	3	69	54	5	2	0	1	4	0	76	0	365	404
16.1. Planctomycetales	0	2	0	0	0	1	6	3	3	14	25	138	41	9	3	69	54	5	2	0	1	4	0	76	0	365	404
16.1.1. Planctomycetaceae	0	2	0	0	0	1	6	3	3	14	25	138	41	9	3	69	54	5	2	0	1	2	0	76	0	365	404
16.1.1.1. Isosphaera	0	0	0	0	0	0	6	3	0	12	25	136	41	9	3	53	50	0	0	0	0	2	0	73	0	353	386
16.1.1.2. Pirellula	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	4	6

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
16.1.1.3. Planctomyces	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0	0	2	0	0	0	0	3	0	0	12
16.1.1.4. Singulisphaera	0	0	0	0	0	1	0	0	3	2	0	0	0	0	0	14	0	5	0	0	1	0	0	0	0	8	0
16.2. Unclassified Planctomycetales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
16. Proteobacteria	1803	5167	14	924	369	1306	352	1786	139	1122	93	1033	3552	1011	2045	7116	1851	5198	67	1696	944	5686	585	3474	13	2384	7689
16.1. Alphaproteobacteria	603	172	4	10	162	399	333	1437	98	1087	26	1030	3181	1005	1388	6289	1525	4910	33	1544	930	4977	378	3020	0	2237	7285
16.1.1. Caulobacterales	13	29	2	0	7	20	1	3	2	0	0	0	0	2	0	4	2	2	0	2	0	19	0	2	0	0	0
16.1.1.1. Caulobacteraceae	13	29	2	0	7	20	1	3	2	0	0	0	0	2	0	4	2	2	0	2	0	19	0	2	0	0	0
16.1.1.1.1. Asticcacaulis	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.1.1.1.2. Brevundimonas	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0
16.1.1.1.3. Caulobacter	0	27	2	0	7	12	0	1	2	0	0	0	0	0	0	3	2	0	0	0	0	0	0	2	0	0	0
16.1.1.1.4. Phenylobacterium	13	0	0	0	0	8	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	19	0	0	0	0	0
16.1.2. Rhizobiales	426	98	2	10	144	118	117	606	58	646	22	529	1155	944	1010	4567	822	3033	30	1194	609	3950	108	2214	13	881	3856
16.1.2.1. Hyphomicrobiaceae	6	7	0	0	0	20	85	210	0	393	0	270	533	581	224	985	230	1004	0	147	3	475	0	311	13	111	578
16.1.2.1.1. Devosia	0	0	0	0	0	4	0	2	0	2	0	7	7	2	8	26	23	7	0	0	0	54	0	7	0	8	26
16.1.2.1.2. Hyphomicrobium	6	7	0	0	0	16	85	208	0	391	0	263	526	579	216	959	207	997	0	147	3	421	0	304	13	83	552
16.1.2.1.3. Rhodomicrobium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0
16.1.2.2. Beijerinckiaceae	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0
16.1.2.2.1. Methylocella	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0
16.1.2.3. Bradyrhizobiaceae	29	1	1	0	0	0	8	9	3	0	0	0	0	0	0	14	0	74	3	0	7	0	108	0	0	0	2
16.1.2.3.1. Agromonas	1	1	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.1.2.3.2. Bradyrhizobium	1	0	0	0	0	0	8	3	0	0	0	0	0	0	0	9	0	73	3	0	4	0	20	0	0	0	2
16.1.2.3.3. Nitrobacter	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
16.1.2.3.4. Oligotropha	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0
16.1.2.3.5. Rhizobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
16.1.2.3.6. Rhodopseudomonas	19	0	0	0	0	0	0	2	3	0	0	0	0	0	0	5	0	0	0	0	0	0	64	0	0	0	0
16.1.2.4. Brucellaceae	3	0	0	0	0	0	0	0	1	0	0	0	2	0	1	0	15	0	0	0	0	1	0	0	0	3	15
16.1.2.4.1. Ochrobactrum	3	0	0	0	0	0	0	0	1	0	0	0	2	0	1	0	15	0	0	0	0	1	0	0	0	3	15
16.1.2.5. Methylobacteriaceae	30	4	0	0	0	26	3	125	18	130	8	136	331	27	27	131	61	153	0	132	45	73	0	165	0	148	392
16.1.2.5.1. Methylobacterium	30	4	0	0	0	26	3	125	18	130	8	136	331	27	27	131	61	153	0	132	45	73	0	165	0	148	392
16.1.2.6. Methylocystaceae	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.1.2.6.1. Methylocystis	4	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.1.2.7. Phyllobacteriaceae	208	57	0	4	14	34	14	68	36	72	10	23	67	139	322	2409	197	205	27	567	281	1488	0	625	0	322	1832
16.1.2.7.1. Chelativorans	0	0	0	0	0	0	0	5	0	19	0	7	33	3	0	5	32	0	0	2	0	4	0	12	0	0	36
16.1.2.7.2. Mesorhizobium	208	57	0	4	14	34	14	63	36	53	10	16	34	136	322	2404	165	205	27	565	281	1484	0	613	0	322	1796
16.1.2.8. Rhizobiaceae	137	26	1	6	130	37	5	172	0	33	4	70	199	181	433	930	283	1413	0	307	268	1841	0	1014	0	251	810
16.1.2.8.1. Agrobacterium	9	6	0	0	0	14	0	37	0	7	0	18	12	12	2	42	48	18	0	0	0	21	0	19	0	15	235

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
16.1.2.8.2. Ensifer	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	1
16.1.2.8.3. Rhizobium	100	14	0	6	42	16	5	95	0	15	4	15	57	126	392	745	77	941	0	290	267	1431	0	783	0	185	372
16.1.2.8.4. Sinorhizobium	28	6	1	0	88	7	0	36	0	11	0	37	130	43	39	143	158	454	0	17	1	388	0	212	0	49	202
16.1.2.9. Xanthobacteraceae	9	3	0	0	0	1	0	20	0	18	0	30	23	16	3	98	36	150	0	41	5	72	0	99	0	46	227
16.1.2.9.1. Unclassified Xanthobacteraceae	9	3	0	0	0	1	0	20	0	18	0	30	23	16	3	98	36	150	0	41	5	72	0	99	0	46	227
16.1.3. Rhodobacterales	122	1	0	0	0	93	12	446	26	164	1	120	707	51	337	1354	249	1639	2	334	315	816	0	292	0	922	1918
16.1.3.1. Rhodobacteraceae	122	1	0	0	0	93	12	446	26	164	1	120	707	51	337	1354	249	1639	2	334	315	816	0	292	0	922	1918
16.1.3.1.1. Dinoroseobacter	13	0	0	0	0	33	0	74	0	2	0	10	226	26	153	409	73	288	0	146	135	309	0	160	0	225	768
16.1.3.1.2. Erythrobacter	0	0	0	0	0	0	1	2	0	3	0	0	0	0	5	22	0	27	0	7	12	8	0	2	0	15	19
16.1.3.1.3. Leisingera	0	0	0	0	0	24	10	133	0	41	0	19	190	18	63	226	33	263	0	101	105	266	0	49	0	128	142
16.1.3.1.4. Paracoccus	106	0	0	0	0	36	1	224	26	110	1	78	262	7	105	683	139	1045	2	75	61	227	0	78	0	531	962
16.1.3.1.5. Rhodobacter	0	1	0	0	0	0	0	12	0	8	0	12	28	0	0	1	3	4	0	0	0	0	0	3	0	2	13
16.1.3.1.6. Roseibacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0
16.1.3.1.7. Roseobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
16.1.3.1.8. Ruegeria	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	11	0	11	0	5	2	4	0	0	0	10	14
16.1.3.1.9. Unclassified Rhodobiaceae	3	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	2	0	0	0	0	0
16.1.4. Rhodospirillales	42	44	0	0	11	165	6	382	2	277	1	381	1318	8	41	363	430	236	1	14	6	186	18	512	0	434	1511
16.1.4.1. Acetobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	18	0	0	0	0
16.1.4.1.1. Acidiphilium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	18	0	0	0	0
16.1.4.2. Rhodospirillaceae	42	44	0	0	11	165	6	382	2	277	1	381	1318	8	41	363	430	233	1	14	6	186	0	512	0	434	1511
16.1.4.2.1. Azospirillum	24	44	0	0	0	156	5	374	2	265	1	362	1265	7	25	274	331	74	0	12	4	80	0	483	0	365	1297
16.1.4.2.2. Magnetospirillum	0	0	0	0	0	1	1	3	0	1	0	0	0	0	0	5	2	2	0	0	0	0	0	2	0	2	5
16.1.4.2.3. Rhodocista	0	0	0	0	0	0	0	2	0	4	0	2	6	1	1	33	12	0	0	2	1	42	0	1	0	20	117
16.1.4.2.4. Rhodospirillum	18	0	0	0	11	7	0	3	0	7	0	17	47	0	12	49	83	118	1	0	1	57	0	26	0	47	89
16.1.4.2.5. Thalassospira	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3	2	2	16	0	0	0	7	0	0	0	0	3
16.1.4.2.6. Tistrella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0
16.1.5. Sphingomonadales	0	0	0	0	0	3	197	0	10	0	2	0	1	0	0	1	22	0	0	0	0	6	0	0	0	0	0
16.1.5.1. Sphingomonadaceae	0	0	0	0	0	3	197	0	10	0	2	0	1	0	0	1	22	0	0	0	0	6	0	0	0	0	0
16.1.5.1.1. Sphingobium	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16.1.5.1.2. Sphingomonas	0	0	0	0	0	0	197	0	10	0	2	0	1	0	0	0	22	0	0	0	0	6	0	0	0	0	0
16.1.5.1.3. Sphingopyxis	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2. Betaproteobacteria	20	3235	0	0	9	3	11	10	5	24	39	0	1	2	22	37	27	144	2	102	6	55	99	128	0	4	74
16.2.1. Burkholderiales	20	1	0	0	0	0	11	1	1	24	39	0	1	2	22	36	22	105	2	100	6	52	29	128	0	2	69
16.2.1.1. Alcaligenaceae	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.1.1.1. Castellaniella	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.1.2. Burkholderiaceae	0	0	0	0	0	0	3	1	0	21	2	0	0	0	21	8	5	98	2	14	3	9	9	4	0	0	3

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
16.2.1.2.1. Acinetobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16.2.1.2.2. ATCC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
16.2.1.2.3. Burkholderia	0	0	0	0	0	0	2	1	0	1	1	0	0	0	21	4	5	98	2	14	3	6	11	0	0	0	3
16.2.1.2.4. Cupriavidus	0	0	0	0	0	0	0	0	0	20	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
16.2.1.2.5. Ralstonia	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
16.2.1.3. Comamonadaceae	14	1	0	0	0	0	8	0	1	0	37	0	1	2	1	22	17	7	0	86	3	38	19	124	0	1	66
16.2.1.3.1. Acidivorax	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	3	0	0	50	0	9	0	2	0	0	0
16.2.1.3.2. Acidovorax	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
16.2.1.3.3. Alicycliphilus	0	0	0	0	0	0	0	0	1	0	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.1.3.4. Comamonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
16.2.1.3.5. Delftia	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0
16.2.1.3.6. Leptothrix	3	0	0	0	0	0	0	0	0	0	0	0	1	2	1	18	6	3	0	36	0	24	0	93	0	1	66
16.2.1.3.7. Polaromonas	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	0	0	0	0	0	0
16.2.1.3.8. Variovorax	1	0	0	0	0	0	8	0	0	0	0	0	0	0	0	2	4	1	0	0	0	3	0	29	0	0	0
16.2.1.4. Oxalobacteraceae	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	3	0	0	0	0	0
16.2.1.4.1. Collimonas	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
16.2.1.4.2. Janthinobacterium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
16.2.1.5. Rubrivivax	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	1	0
16.2.1.5.1. Rubrivivax	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	1	0
16.2.1.6. Burkholderiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
16.2.1.6.1. Burkholderiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
16.2.2. Gallionellales	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.2.1. Gallionella	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.2.1.1. Siderooxidans	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.3. Neisseriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
16.2.3.1. Chromobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
16.2.3.1.1. Pseudogulbenkiania	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
16.2.4. Nitrosomonadales	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0
16.2.4.1. Nitrosomonadaceae	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0
16.2.4.1.1. Nitrosomonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0
16.2.4.1.2. Nitrosospira	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.2.5. Rhodocyclales	0	0	0	0	0	0	0	9	4	0	0	0	0	0	0	1	2	5	0	0	0	3	0	0	0	2	5
16.2.5.1. Rhodocyclaceae	0	0	0	0	0	0	0	9	4	0	0	0	0	0	0	1	2	5	0	0	0	3	0	0	0	2	5
16.2.5.1.1. Aromatoleum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	3	0	0	0	2	0
16.2.5.1.2. Azoarcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
16.2.5.1.3. Thauera	0	0	0	0	0	0	0	9	4	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
16.2.6. Sulfuricellales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
16.2.6.1. Sulfuricellaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
16.2.6.1.1. Sulfuricella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
16.2.7. Unclassified Betaproteobacteria	0	3234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
16.2.7.1. Candidatus	0	3234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
16.2.7.1.1. Candidatus-Accumulibacter	0	3234	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
16.3. Deltaproteobacteria	2	1	3	0	197	13	2	6	1	7	0	3	0	0	0	8	2	1	0	13	0	0	0	3	0	4	2
16.3.1. Desulfuromonadales	0	1	3	0	197	1	0	2	1	3	0	0	0	0	0	5	1	0	0	13	0	0	0	2	0	4	0
16.3.1.1. Geobacteraceae	0	0	0	0	196	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0
16.3.1.1.1. Geobacter	0	0	0	0	196	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0
16.3.1.2. Pelobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
16.3.1.2.1. Pelobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
16.3.1.3. Anaeromyxobacteraceae	0	1	0	0	0	1	0	0	0	2	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	4	0
16.3.1.3.1. Anaeromyxobacter	0	1	0	0	0	1	0	0	0	2	0	0	0	0	0	1	1	0	0	2	0	0	0	0	0	4	0
16.3.1.4. Archangiaceae	0	0	3	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0
16.3.1.4.1. Stigmatella	0	0	3	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0
16.3.1.5. Haliangiaceae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.3.1.5.1. Haliangium	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.3.2. Myxococcales	2	0	0	0	0	12	2	4	0	4	0	3	0	0	0	3	1	1	0	0	0	0	0	1	0	0	2
16.3.2.1. Myxococcaceae	0	0	0	0	0	12	0	0	0	4	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	2
16.3.2.1.1. Chondrococcus	0	0	0	0	0	11	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.3.2.1.2. Myxococcus	0	0	0	0	0	1	0	0	0	2	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	2
16.3.2.2. Sorangiaceae	2	0	0	0	0	0	2	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16.3.2.2.1. Sorangium	2	0	0	0	0	0	2	4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16.3.2.3. Syntrophobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
16.3.2.3.1. Syntrophobacter	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
16.4. Gammaproteobacteria	1178	1759	7	914	1	891	6	333	35	4	28	0	370	4	635	782	297	143	32	37	8	654	108	323	0	139	328
16.4.1. Aeromonadales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	27	0	0	0	3	0	0	0	0	8
16.4.1.1. Aeromonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	27	0	0	0	1	0	0	0	0	8
16.4.1.1.1. Aeromonas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	27	0	0	0	1	0	0	0	0	8
16.4.1.2. Alteromonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0
16.4.1.2.1. Marinobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
16.4.1.2.2. Microbulbifer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
16.4.2. Chromatiales	0	0	0	1	0	3	0	0	0	1	0	0	0	0	0	0	0	47	8	0	0	0	8	1	0	0	2
16.4.2.1. Chromatiaceae	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	47	8	0	0	0	8	1	0	0	0
16.4.2.1.1. Allochromatium	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	8	0	0	0	8	0	0	0	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
16.4.2.1.2. Nitrosococcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
16.4.2.1.3. Thioflavicoccus	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4.2.2. Ectothiorhodospira	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
16.4.2.2.1. Alkalilimnicola	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4.2.2.2. Thioalkalivibrio	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
16.4.3. Enterobacteriales	13	0	0	0	0	12	0	48	14	0	22	0	4	4	0	17	4	21	0	0	0	20	0	17	0	1	5
16.4.3.1. Enterobacteraceae	13	0	0	0	0	12	0	48	14	0	22	0	4	4	0	17	4	21	0	0	0	20	0	17	0	1	5
16.4.3.1.1. Aerobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	1	0	0	0	1	0	0	0	1	1
16.4.3.1.2. Unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16.4.3.1.3. Erwinia	0	0	0	0	0	0	0	0	5	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4.3.1.4. Pantoea	0	0	0	0	0	12	0	48	9	0	0	0	0	2	0	2	3	0	0	0	0	0	0	0	0	0	4
16.4.3.1.5. Serratia	13	0	0	0	0	0	0	0	0	0	0	0	4	2	0	9	0	20	0	0	0	19	0	17	0	0	0
16.4.4. Oceanospirillales	0	0	0	0	0	0	0	22	3	0	5	0	0	0	5	13	0	0	0	0	0	0	0	0	0	0	0
16.4.4.1. Halomonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
16.4.4.1.1. Chromohalobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
16.4.4.2. Hahellaceae	0	0	0	0	0	0	0	22	3	0	5	0	0	0	5	8	0	0	0	0	0	0	0	0	0	0	0
16.4.4.2.1. Hahella	0	0	0	0	0	0	0	22	3	0	5	0	0	0	5	8	0	0	0	0	0	0	0	0	0	0	0
16.4.5. Pseudomonadales	1159	1759	7	913	0	876	4	256	18	3	0	0	366	0	630	750	290	48	24	37	8	631	80	305	0	138	313
16.4.5.1. Pseudomonadaceae	1159	1759	7	913	0	876	4	256	18	3	0	0	366	0	630	750	290	48	24	37	8	631	80	305	0	138	313
16.4.5.1.1. Azotobacter	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16.4.5.1.2. Chlorobacterium	23	0	1	627	0	64	0	26	5	0	0	0	0	0	13	0	5	0	1	0	0	0	0	0	0	0	0
16.4.5.1.3. Pseudomonas	1134	1759	6	286	0	812	4	230	13	3	0	0	366	0	617	749	285	48	23	37	8	631	80	305	0	138	313
16.4.6. Xanthomonadales	6	0	0	0	1	0	2	7	0	0	1	0	0	0	0	0	1	0	0	0	0	0	20	0	0	0	0
16.4.6.1. Lysobacteraceae	6	0	0	0	1	0	2	7	0	0	1	0	0	0	0	0	1	0	0	0	0	0	20	0	0	0	0
16.4.6.1.1. Phytomonas	0	0	0	0	1	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4.6.1.2. Pseudoxanthomonas	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16.4.6.1.3. Rhodanobacter	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
16.4.6.1.4. Stenotrophomonas	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
17. Spirochaetae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
17.1. Spirochaetales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
17.1.1. Leptospiraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
17.1.1.1. Leptospira	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
18. Verrucomicrobia	8	2	0	0	4	0	0	4	7	3	1	0	0	0	35	5	5	6	3	83	11	15	0	22	21	1	0
18.1. Opitutales	8	2	0	0	4	0	0	4	0	3	1	0	0	0	35	3	5	6	3	83	11	15	0	20	21	1	0

Table S4.5. Taxonomic summary of the *phoX* gene showing the number of filtered reads at the phylum, class, order, family and genus levels in the samples. (Continued)

Taxonomy	AUS1-A	AUS1-GF	AUS2-A	AUS2-Gf	AUS2-Guf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS4-F	AUS4-G	AUS5-F	AUS5-G	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-A	CH5-F	CH5-G
18.1.1. Opitutaceae	8	2	0	0	4	0	0	4	0	3	1	0	0	0	35	3	5	6	3	83	11	15	0	20	21	1	0
18.1.1.1. Opitutus	8	0	0	0	4	0	0	4	0	3	0	0	0	0	0	3	5	6	0	0	0	15	0	9	21	0	0
18.1.1.2. Unclassified Opitutaceae	0	2	0	0	0	0	0	0	0	0	1	0	0	0	35	0	0	0	3	83	11	0	0	11	0	1	0
18.2. Puniceicoccales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
18.2.1. Puniceicoccaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
18.2.1.1. Coraliomargarita	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
18.3. Verrucomicrobiales	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
18.3.1. Verrucomicrobia	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
18.3.1.1. Akkermansia	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
19. Unclassified	7	7	0	19	7	6	5	6	3	25	5	7	0	0	32	19	32	37	3	36	1	14	550	27	13	38	0

5

MICROBIAL COMMUNITIES AND PHOD AS AFFECTED BY PHOSPHATE DEPLETION AND PH IN SOIL

Accepted as:

Ragot SA, O Huguenin-Elie, MA Kertesz, E Frossard and EK Bünemann (2016). Microbial communities and *phoD* as affected by phosphate depletion and pH in soil. *Plant and Soil*. DOI: 10.1007/s11104-016-2902-5.

Abstract

Soil microbial communities contribute to organic phosphorus cycling in a variety of ways, including secretion of the PhoD alkaline phosphatase. We sampled a long-term grassland fertilization trial in Switzerland characterized by a natural pH gradient. We examined the effects of phosphate depletion and pH on total and active microbial community structures and on the structure and composition of the total and active *phoD*-harboring community.

Archaeal, bacterial and fungal communities were investigated using T-RFLP and *phoD*-harboring members of these communities were identified by 454-sequencing.

Phosphate depletion decreased total, resin-extractable and organic phosphorus in soil and changed the structure of all active microbial communities, and of the total archaeal and *phoD*-harboring communities. Organic carbon, nitrogen and phosphorus increased with pH, and the structures of all total and active microbial communities except the total fungal community differed between the two pH levels. *phoD*-harboring members were affiliated to *Actinomycetales*, *Bacilliales*, *Gloeobacterales*, *Planctomycetales* and *Rhizobiales*.

Our results suggest that pH and associated soil factors are important determinants of microbial and *phoD*-harboring community structures. These associated factors include organic carbon and total nitrogen, and to a lesser degree phosphorus status, and active communities are more responsive than total communities. Key players in organic phosphorus mineralization are affiliated to phyla that are known to be important in organic matter decomposition.

5.1 Introduction

Soils sustain complex and diverse microbial communities that are intimately associated with their habitat. Soil properties and vegetation fundamentally shape the microbial community by affecting its composition, diversity and functioning (Berg and Smalla 2009; Dequiedt et al. 2011). Within the soil microbial community, archaea, bacteria and fungi are influenced differently by soil properties and by the plants present. Among the soil properties, soil pH has repeatedly been shown to be the primary environmental driver influencing the total bacterial community structure studied at DNA level (Fierer and Jackson 2006; Lauber et al. 2008; Griffiths et al. 2011). The total fungal community has been reported to be less responsive to pH changes than the total bacterial community (Blagodatskaya and Anderson 1998; Wakelin et al. 2008), and only some specific groups of archaea such as ammonia oxidizers have been shown to change with pH (Nicol et al. 2008). Specific taxonomic groups such as the bacterial groups *Acidobacteria, Bacteroidetes* and *Firmicutes*, and the fungal *Helotiales, Hypocreales* and mitosporic *Basidomycetes* have been reported to be highly responsive to pH (Lauber et al. 2009; Rousk et al. 2010). However, the effect of pH on the active community of these microorganisms (studied at RNA level) remains unclear.

Soil pH also affects the availability of nutrients such as phosphorus (P) by modifying adsorption and desorption reactions (Frossard et al. 1995). This makes it difficult to distinguish the effect of P availability on microbial communities from that of pH. P is often a limiting nutrient for plants and microorganisms in soil (Vitousek et al. 2010) and phosphate fertilizer is therefore commonly applied to agricultural soils to maintain crop productivity. Phosphate addition can lead to a shift in the microbial community composition (Mander et al. 2012; Siciliano et al. 2014). Archaea and bacteria are generally thought to be more responsive than fungi to water-soluble phosphate addition (Cruz et al. 2009; Bissett et al. 2011). Under low P availability, microorganisms up-regulate the genes of the Pho regulon, which includes genes encoding phosphatases, phosphate

transporters and other systems for mobilizing soil P (Vershinina and Znamenskaya 2002). Phosphatases can potentially hydrolyze up to 89 % of the total extracted organic P extracted from soil (Jarosch et al. 2015).

The *phoD* alkaline phosphatase is found in a range of environments (Luo et al. 2009; Tan et al. 2013) and is mainly produced by bacteria (Chapter 2). It hydrolyzes phosphomonoesters and –diesters excluding inositol phosphates such as phytate Kageyama et al. (2011), which represent an important fraction of organic P in soil (Condron et al. 2005). The few studies addressing the diversity, total community structure and composition of the *phoD* gene report effects of organic and water-soluble phosphate fertilizers, crop management, vegetation and pH (Sakurai et al. 2008; Wang et al. 2012a; Chhabra et al. 2013; Jorquera et al. 2014; Tan et al. 2013; Ragot et al. 2015). Effects of soil pH on the total *phoD*-harboring community have been reported in studies in which organic C and available P increased together with pH (Wang et al. 2012b; Cui et al. 2015). Using pyro-sequencing, dominant *phoD*-harboring community members have been shown to be affiliated primarily to *Alpha-* and *Gamma-Proteobacteria* but also to *Acidobacteria* and *Cyanobacteria* (Tan et al. 2013; Fraser et al. 2015b). However, these studies used a set of primers (ALPS primers; Sakurai et al. (2008)) that has been shown to have an amplification bias towards *Alpha-Proteobacteria* and to cover the *phoD* diversity only partly (Tan et al. 2013; Ragot et al. 2015).

This study aimed to clarify the effect of phosphate depletion and soil pH on microbial communities in a long-term fertilization trial on a permanent grassland characterized by a pH gradient and low P availability (Bünemann et al. 2012). To evaluate the effect of phosphate depletion, we compared microbial communities in plots fertilized with N, P and K, which maintained soil P stocks, with plots fertilized with N and K only, which resulted in phosphate depletion (Liebisch et al. 2013). To evaluate the effect of pH, we sampled plots immediately adjacent to the fertilization trial, which had two different pH values.The total and active archaeal, bacterial, fungal and *phoD*-harboring community structures were examined, and related to soil and plant properties using multivariate

analysis, allowing us to separate the effects of phosphate depletion and pH. In addition, we analyzed the composition of the total and active *phoD*-harboring communities to identify the key microorganisms that harbor and/or express the *phoD* gene. It has been shown that the total and active microbial community can differ significantly during decomposition processes, as only some members of the total community (especially *Acidobacteria, Actinobacteria* and *Proteobacteria*) are metabolically active (Baldrian et al. 2012). We therefore expected stronger expression of the *phoD* gene by these taxonomic groups than by other *phoD*-harboring organisms, and the active community structures to differ more between treatments than the total community structures.

5.2 Material and Methods

5.2.1 Site description and sampling

Soils were sampled in September 2012 from a long-term fertilization trial (1992-2014) at the research station Agroscope near Watt in the Swiss lowland (47°26'44"N, 8°29'32"E, 500 m asl) (Huguenin-Elie et al. 2006). The vegetation consists of a permanent grassland of *Arrhenatherion elatioris* association with 30 to 40 angiosperm species (Liebisch et al. 2013), which is harvested as hay three times per year with a late first harvest (after June 15th) to preserve plant biodiversity. The mesic grassland grows on a medium-deep (50-70 cm) Cambisol (IUSS 2014) with a medium base saturation. The soil in the upper 10 cm consists of 220, 340 and 440 g kg^{-1} clay, silt and sand, respectively (Philipp et al. 2004).

The fertilization trial has a randomized block design with four replicates and a plot size of 2 m x 5 m (Figure 5.1). The treatments represent different combinations of mineral N, P and K additions. In this study, we selected two treatments: one fertilized with N, P and K (NPK) and one fertilized with N and K only (NK). The amounts added were 45 kg N ha^{-1} yr^{-1} as ammonium nitrate, 17 kg P ha^{-1} yr^{-1} as single superphosphate and 83 kg K ha^{-1} yr^{-1} as potassium chloride. The phosphate input in NPK plots was adequate to reach

a zero P balance, while phosphate depletion via the negative P balance of -10 kg P ha^{-1} yr^{-1} in NK decreased P availability (Liebisch et al. 2013).

Previous work indicated the presence of a pH gradient along the field that affects mainly one NPK plot (Southwest corner plot in block IV), which has a higher pH than the other NK and NPK plots (Liebisch et al. 2013). In order to assess the effect of pH, an additional set of samples was taken in the border strip next to the original field trial at both the lower (pH 5.0) and the higher (pH 6.1) pH end of the trial by extending each block by one plot on each side as indicated in Figure 5.1. This border strip has not been fertilized since 1992, but the vegetation has been cut and removed along with harvests in the trial.

Five soil cores (0-5 cm) were randomly collected in each plot, combined and homogenized by sieving through an 8 mm sieve directly after sampling. A subsample of soil was immediately frozen in liquid nitrogen in the field and stored at -80°C until molecular analysis. Further subsamples were stored at -20°C for later enzymatic analysis, or dried and ground for determination of total nutrient contents. The remaining soil was stored at 4°C for 2 weeks until analysis.

5.2.2 Soil and vegetation analyses

5.2.2.1 Total C, N and P, and pH

Total organic C (TOC) and total N (TN) were measured on finely ground soil using a CNS analyzer (Thermo-Fisher Flash EA 1112). Total P (TP) in soil was determined by wet digestion with H_2O_2/H_2SO_4 (Anderson and Ingram 1993) and measured with malachite green at 610 nm using a UV-1800 Shimadzu Spectrometer (Ohno and Zibilske 1991). Soil pH was measured in a soil suspension in 0.01M $CaCl_2$ (1 to 2.5 soil mass to volume ratio, shaken for 1 h) using a Benchtop pH 720A (Orion Research Inc., Jacksonville, FL).

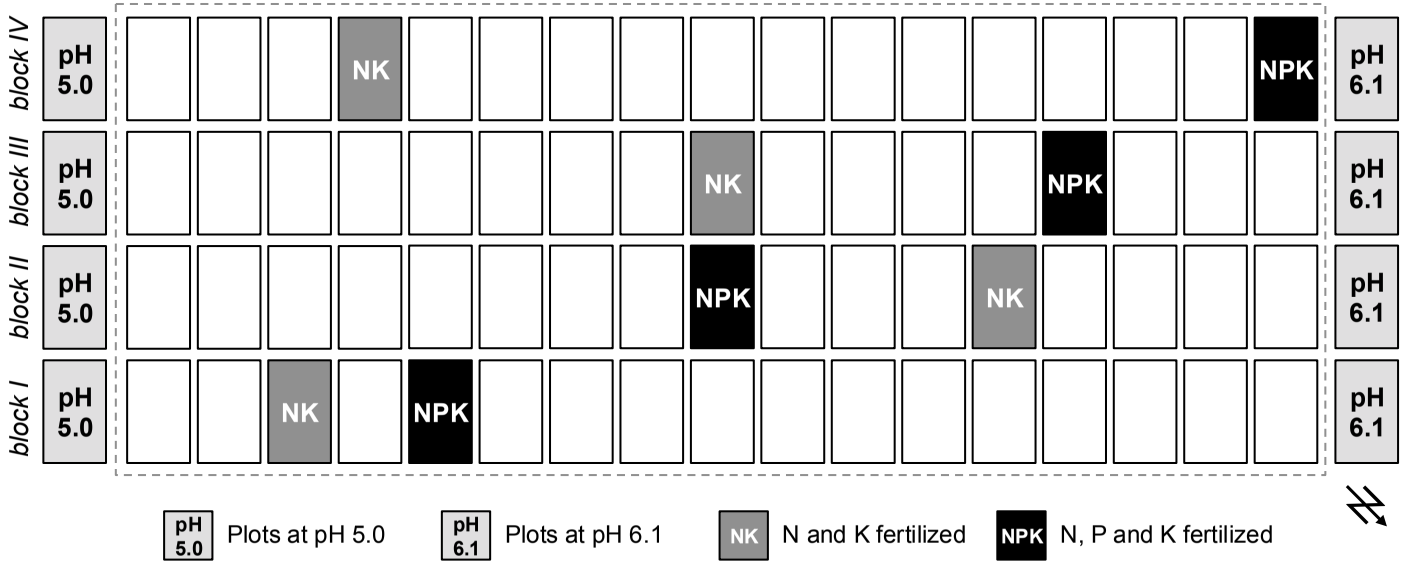


Figure 5.1. Layout of the long-term field trial in Switzerland showing the selected plots in the fertilization trial (NK and NPK treatments) and the additional pH levels.

5.2.2.2 NaOH-EDTA-extractable organic P

Organic P was determined using alkaline extraction after Bowman and Moir (1993). Briefly, 2.5 g of finely ground soil was extracted with 0.25 M NaOH - 0.05 M EDTA in a 1 to 10 (w/v) ratio for 16 h on a horizontal shaker. Soil extracts were then centrifuged and filtered (Whatman No. 1). The concentration of inorganic P in the extract was measured colorimetrically using malachite green (Ohno and Zibilske 1991). The concentration of total P in the extract was measured colorimetrically using malachite green on neutralized extracts after digestion with ammonium persulphate in an autoclave (20 min, 121°C). Organic P was calculated by subtracting inorganic P from total P in the extract.

5.2.2.3 Microbial C, N and P and resin-extractable P

Microbial C and N analysis was done using chloroform fumigation-extraction on 25 g of moist soil (Vance et al. 1987) . Total organic C and total N in the extracts were measured using a Formacs^{SERIES} TOC/TN analyzer (Skalar, The Netherlands). Concentrations in non-fumigated subsamples were subtracted from those in fumigated ones to calculate chloroform-labile C (C_{mic}) and N (N_{mic}). No conversion factor was applied since this is soil-specific and has not been determined for the soils in this study.

Microbial and resin-extractable P were determined on moist samples by fumigation-extraction using anion-exchange resin membranes (BDH laboratory supplies #55164 2S, England) saturated with CO_3^{2-} and 1-hexanol as fumigant (Kouno et al. (1995) as modified by Bünemann et al. (2007)). Microbial P (P_{mic}) was calculated as the difference between fumigated and non-fumigated subsamples (resin-extractable P, P_{res}) and corrected for the incomplete recovery of a P spike, which ranged between 88% and 99%. No conversion factor was applied, for the same reason as above.

5.2.2.4 Potential acid and alkaline phosphatase activity

Potential acid and alkaline phosphatase activity were determined following Marx et al. (2001) as modified by Poll et al. (2006). Briefly, one gram of moist soil was dispersed in

100 mL of autoclaved ddH_2O using an ultrasonic probe Labsonic U (Braun Melsungen, Germany) at 50 J s^{-1} for 120 s. The assay contained 50 µL aliquots of the soil suspension in a microplate with 6 analytical replicates, using 1 mM 4-methylumbelliferylphosphate as substrate and either 0.1 M MES buffer (pH 6.1) or Modified Universal Buffer (pH 11) (Alef and Nannipieri 1995) as buffer for acid and alkaline phosphatase activity, respectively. Microplates were incubated on a horizontal shaker for 3 h at 30°C. The linear increase in fluorescence over time was measured, and the enzymatic activity (µmol h^{-1} g $soil^{-1}$) was calculated by normalization to a 4-methylumbelliferone standard added to the soil suspension.

5.2.2.5 Vegetation

All vascular plant species were identified and their relative abundance at heading stage of the dominant grass species was visually estimated in May 2013 following Dietl (1995). Plant samples for analysis of P and N content were taken in June 2013, shortly before the first cut of the year. The vegetation was cut at 4 cm above the ground using electric scissors in a randomly selected 50 x 50 cm square with at least 0.5 m distance to the border of the plot. The material was sorted into grasses, legumes and forbs, and the three groups were dried at 60°C for 3 days and milled. Plant N content was measured by a CNS analyzer. To determine plant P content, 250 mg of plant material was ashed (6 h, 550°C) and dissolved in 3 mL 10.3 M HNO_3. Samples were diluted and neutralized before colorimetric P determination using malachite green.

5.2.3 Molecular analysis

5.2.3.1 Nucleic acid extraction and reverse transcription

Nucleic acids were extracted in analytical duplicates from 2 g of snap-frozen soil using the RNA Powersoil® Total RNA isolation kit (MO BIO, Carlsbad, CA, USA) according to the manufacturer's instructions, with an additional homogenizing step (2.8 mm zirconium

beads for 1 min at 5 m s^{-1}) using the Omni Bead Ruptor Homogenizer (Omni International, Kennesaw, GA) prior to isolation. RNA was eluted from the RNA/DNA Capture Column, precipitated and resuspended following the manufacturer's instructions. Subsequently, DNA was eluted using 4 mL of DNA elution solution (1 M NaCl, 50 mM MOPS, 15 % (v/v) isopropanol, pH 7). DNA was precipitated with isopropanol, washed, dried, and redissolved in DEPC-treated H_2O. RNA samples were treated with Turbo DNase (Thermo Fisher Scientific Inc., Waltham, MA) to remove DNA contamination. RNA concentrations were measured using a Qubit® 2.0 Fluorometer (Life Technologies, Inc., Grand Islands, NY). Complementary DNA (cDNA) was synthesized from 200 ng of total RNA, using 500 ng random hexamers (Thermo Fisher Scientific Inc.) and 160 Units GoScript™ Reverse Transcriptase (Promega, Madison, WI) according to the manufacturer's recommendations.

5.2.3.2 Analysis of the archaeal, bacterial and fungal community structures using T-RFLP

Archaeal, bacterial and fungal communities were studied using regions of the 16S rRNA gene for archaea and bacteria, and the ITS region for fungi, as detailed below. All PCRs (25 µL) contained 0.6 Units GoTaq (Promega), 1X reaction GoTaq Buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer, and 1-2 ng template DNA. Amplification of archaeal 16S DNA and cDNA was carried out with Ar109F/FAM-Ar915R primers (Lueders and Friedrich 2000). Bacterial 16S DNA and cDNA were amplified using 1064R/FAM-356F primers (Winsley et al. 2012) . The fungal community was investigated with FAM-ITS1-F (Gardes and Bruns 1993)/ITS4-R (White et al. 1990) primers. After initial denaturation for 5 min at 95°C, amplification was done for 35 cycles comprising a denaturation step of 30 s at 95°C, an annealing step of 30 s at 53°C (archaea), 60 s at 52°C (bacteria) or 30 s at 60°C (fungi), and an extension step of 60 s at 72°C. A final extension step of 5 min at 72°C was carried out to complete the reaction. Results

from DNA- and cDNA-analysis are referred to as measurements of the total and active communities, respectively.

Restriction enzyme digestion reaction contained 10 µL PCR product, 5 U of *MspI* (Promega), 2 µg acetylated bovine serum albumin and water to a final volume of 20 µL, and was carried out for 3 h at 37°C. Digested products were precipitated by adding 3 µL of 3 M CH_3COONa (pH 5.2), 3 µL of 100 mM Na_2EDTA pH 8.0), 400 µg of glycogen and 60 µL 95 % (v/v) ice-cold ethanol. Products were centrifuged, washed with 70 % (v/v) ice-cold ethanol and redissolved in 15 µL DEPC-treated H_2O. The purified products (3 µL) were analyzed on an ABI 3130XL sequencer using Liz^{500} as standard (Life Technologies, Inc.). The T-RFLP electropherograms of the archaeal, bacterial and fungal communities were analyzed in GeneMapper v3.7 (Applied Biosystems, Foster City, CA) with a fragment length range of 50-500 base pairs. T-RFLP profiles were processed and analyzed with the T-REX program (Culman et al. 2009), using the peak area to identify operational taxonomic units (OTUs) and a clustering threshold of 0.5 bp.

5.2.3.3 Analysis of total and active *phoD*-harboring community structure and composition using 454-sequencing

The *phoD* alkaline phosphatase gene was amplified by PCR from DNA and cDNA using *phoD*-F733/R1083 primers (See Chapter 2) in a reaction mixture (25 µL) containing 0.6 Units GoTaq (Promega) 1X GoTaq buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM of each primer, and 1-2 ng DNA or cDNA as template. Amplification included an initial denaturation step of 5 min at 95°C, followed by 35 cycles of a denaturation step of 30 s at 95°C, an annealing step of 30 s at 58°C and an extension step of 60 s at 72°C. The final extension step was carried out for 10 min at 72°C. Field replicates of DNA and cDNA extracts of each plot were diluted to 20 ng μL^{-1}, pooled and analyzed by 454-sequencing using a GS-FLX+ platform (Roche 454 Life Sciences, Branford, CT), at Research and Testing Laboratory (Lubbock, TX). The Standard Flowgram Format (SFF)

files were submitted to the European Nucleotide Archive (ENA) under the accession number ERP010357.

Reads resulting from the 454-sequencing were analyzed using MOTHUR (Schloss et al. 2004) as described in section 2.2.7. Briefly, resulting reads were trimmed at 150 bp and 450 bp as minimum and maximum length, respectively, and aligned using the Needleman-Wunsch global alignment algorithm as implemented in MOTHUR, using 6-mers searching and aligned reference sequences as template as in section 2.2.7. The pairwise distance matrix was calculated from the alignment and sequences clustered using the k-furthest method as implemented in MOTHUR with a similarity cutoff at 75% to define the operational taxonomic units (OTUs) as calculated by Tan et al. (2013). OTU matrices were then normalized to the smallest library size using the normalized.shared command as implemented in MOTHUR to allow comparison between plots (Table S5.5). Taxonomy was assigned using BLAST+ (Camacho et al. 2009) with a minimum e-value of 1e-8 to retrieve NCBI sequence identifiers (GI accession number), and in-house Perl scripts were then used to obtain the taxonomy for each read (scripts written by Stefan Zoller, Genetic Diversity Centre, ETH Zürich) (Table S5.5).

5.2.4 Statistical analysis

Since the plots with different pH levels were not part of the randomized block design of the fertilization trial, soil properties and plant data from the fertilization trial and from the two pH levels were subjected to independent statistical analysis. Redundancy analysis (RDA) was then used to evaluate the effect of soil and plant properties on microbial community structures across the whole dataset.

Two-tailed *t*-test was used to examine differences between fertilization treatments, and between pH levels, respectively. Analysis of dissimilarity (*anosim*) in the vegan package (vegan: Community Ecology Package) in **R** v.2.15.0 (**R** Core Team, http://www.R-project.org/, 2014) was used to test whether the vegetation composition was significantly

different between fertilization treatments and between pH levels, respectively. Linear correlations between soil properties were tested in a Pearson's correlation matrix (Table S5.2).

T-RFs of the archaeal, bacteria and fungal communities were normalized using Hellinger distance (Legendre and Gallagher 2001) and T-RFs were further converted into a binary matrix. Species richness was represented by the number of OTUs found in the T-RF profiles in each community (Table S5.3). Analysis of dissimilarity was used to evaluate whether the T-RF profiles of the total and active microbial communities were significantly different between fertilization treatments and between pH levels.

Rarefaction curves based on *phoD* reads were calculated and extrapolated to an average library size of 5,099 reads to standardize the samples using EstimateS (Version 9, http://purl.oclc.org/estimates). Species richness of the *phoD*-harboring community was calculated using the Chao1 index (Chao and Shen 2003) (Table S5.3). Structures of the *phoD*-harboring community within the fertilization treatments and within the pH levels were compared using pairwise *Libshuff* analysis with 1000 iterations as implemented in MOTHUR (Schloss et al. 2004).

Relationships between the microbial community structures and soil and plant properties standardized by the z-score method (Ramette 2007) were analyzed using redundancy analysis (RDA) on T-RFLP data for the archaeal, bacterial and fungal community and on 454-sequencing data for the *phoD*-harboring community. Plant species were only included in the analysis if they had an abundance of more than 2 % and differed significantly in abundance ($p < 0.05$) in either fertilization treatments or pH levels (Table S5.1).

Table 5.1. General soil physicochemical and biological properties (means of four field replicates ± standard deviation). Letters indicate significant differences within the fertilization treatments and within the pH levels, respectively (p <0.05).

Plots	pH	TOC	TN	TP	P_{org}	P_{res}	C_{mic}	N_{mic}	P_{mic}	Acid phosphatase activity	Alkaline phosphatase activity
	($CaCl_2$)	(g kg^{-1})		(mg kg^{-1})						(nmol substrate g^{-1} h^{-1})	
NK	5.1 ± 0.3	31.2 ± 1.1	2.88 ± 0.17	610 ± 35^{a}	336 ± 18^{a}	0.3 ± 0.1^{a}	432 ± 13	86.8 ± 4.5	37.4 ± 6.8^{a}	5113 ± 543^{a}	196 ± 50
NPK	5.3 ± 0.3	30.7 ± 2.4	2.85 ± 0.16	720 ± 48^{b}	372 ± 13^{b}	2.5 ± 1.2b	510 ± 75	96.6 ± 3.8	50.2 ± 4.1^{b}	3434 ± 647^{b}	221 ± 73
pH 5.0	5 ± 0.1^{a}	27.5 ± 0.3^{a}	2.65 ± 0.09^{a}	612 ± 32^{a}	331 ± 17	0.3 ± 0.1	398 ± 14	91.3 ± 8.2	39.7 ± 2.9^{a}	4108 ± 643^{a}	225 ± 61^{a}
pH 6.1	6.1 ± 0.3^{b}	34.5 ± 2.1^{b}	3.41 ± 0.31^{b}	703 ± 31^{b}	320 ± 20	0.2 ± 0.1	383 ± 23	103.9 ± 12.5	54.8 ± 3.8^{b}	3037 ± 305^{b}	1150 ± 325^{b}

5.3 Results

5.3.1 General soil physicochemical and biological properties

Soil pH was similar in the two fertilization treatments, although one NPK replicate had a pH of 5.9, while pH in all other NK and NPK plots varied between 4.9 and 5.3 (Table 5.1). Phosphate depletion in NK decreased TP, P_{org}, P_{res}, P_{mic} and acid phosphatase activity compared to NPK. The pH levels differed significantly by one pH unit. At higher pH values, concentrations of TOC, TN and TP were greater than at lower pH.

Potential acid and alkaline phosphatase activities were strongly correlated with pH, with alkaline phosphatase activity being five times greater at pH 6.1 than at pH 5.0, whereas acid phosphatase activity showed the opposite trend (Table S5.2). Among the other soil properties, pH was linearly correlated with TOC, TN, TP, N_{mic}, N_{mic} and P_{mic} (Table S5.2).

5.3.2 Plant productivity and composition

Phosphate depletion led to a 1.8-fold decrease in total plant yield, though different plant groups were differently affected, with a 2.7-fold decrease in grass biomass contrasting with a 2-fold increase in forb biomass (Table 5.2). Phosphate depletion also decreased plant N and P content by factors of two and four, respectively. In total, 40 plant species were observed in the fertilization treatments (Table S5.1). The vegetation composition was significantly different between the fertilization treatments, while the plant richness remained similar (p-value = 0.03). Among the eight plant species differing in abundance between NK and NPK treatments, the legumes *Trifolium pratense* and *Lotus corniculatus* were more abundant in NK than NPK, while the grasses *Arrhenatherum elatius* and *Holcus lanatus* were more abundant in NPK than NK.

Grass, legume and forb yields were similar at both pH levels (Table 5.2). Lower pH led to a higher plant N content, associated with a slightly reduced species richness. A total of 42 plant species was observed at the two pH levels (Table S5.1). Vegetation composition was significantly different between the pH levels (p-value = 0.03). Acidomesophilic plant species such as *Anthoxanthum odoratum, Cynosurus cristatus* and *Rumex* acetosa were more abundant at pH 5.0, while alkalomesophilic plant species such as *Knautia arvensis* and *Galium album* were more abundant at pH 6.1. This shows that one unit difference in pH can significantly influence the vegetation composition.

Table 5.2. Plant productivity, plant N and P content of the first cut in 2013 (means of four field replicates ± standard deviation) and plant species richness. Letters indicate significant differences between fertilization treatments and pH levels (p<0.05).

Plots	Grass yield T ha^{-1}	Legume yield T ha^{-1}	Forb yield T ha^{-1}	Total yield T ha^{-1}	Plant N content kg ha^{-1}	Plant P content kg ha^{-1}	Plant species richness
NK	1.8 ± 0.2^{a}	0.2 ± 0.2	1.0 ± 0.5	3 ± 0.3^{a}	38.2 ± 1.5^{a}	2.46 ± 0.53^{a}	33 ± 3
NPK	4.9 ± 0.8^{b}	0.2 ± 0.1	0.4 ± 0.1	5.5 ± 0.9^{b}	65.4 ± 7.3^{b}	8.71 ± 0.39^{b}	33 ± 3
pH 5.0	1.7 ± 0.5	0.3 ± 0.2	1.1 ± 0.5	3.1 ± 0.6	56.5 ± 9.8^{a}	2.83 ± 0.58	28 ± 3^{a}
pH 6.1	1.1 ± 0.2	0.2 ± 0.1	1.2 ± 0.1	2.5 ± 0.4	36.6 ± 7.3^{b}	2.12 ± 0.54	34 ± 2^{b}

5.3.3 Effect of phosphate depletion and soil pH on total and active archaeal, bacterial, fungal and *phoD*-harboring community structures

Phosphate depletion led to a significant shift in both total and active archaeal and *phoD*-harboring community structures (Table 5.3). For bacteria and fungi, an effect of phosphate depletion on active but not on total community structures was observed.

The difference in pH of about one unit between pH levels affected the structure of both total and active communities of archaea, bacteria and *phoD* (Table 5.3). The total fungal community structure was similar at both pH values, whereas the active fungal community structure differed.

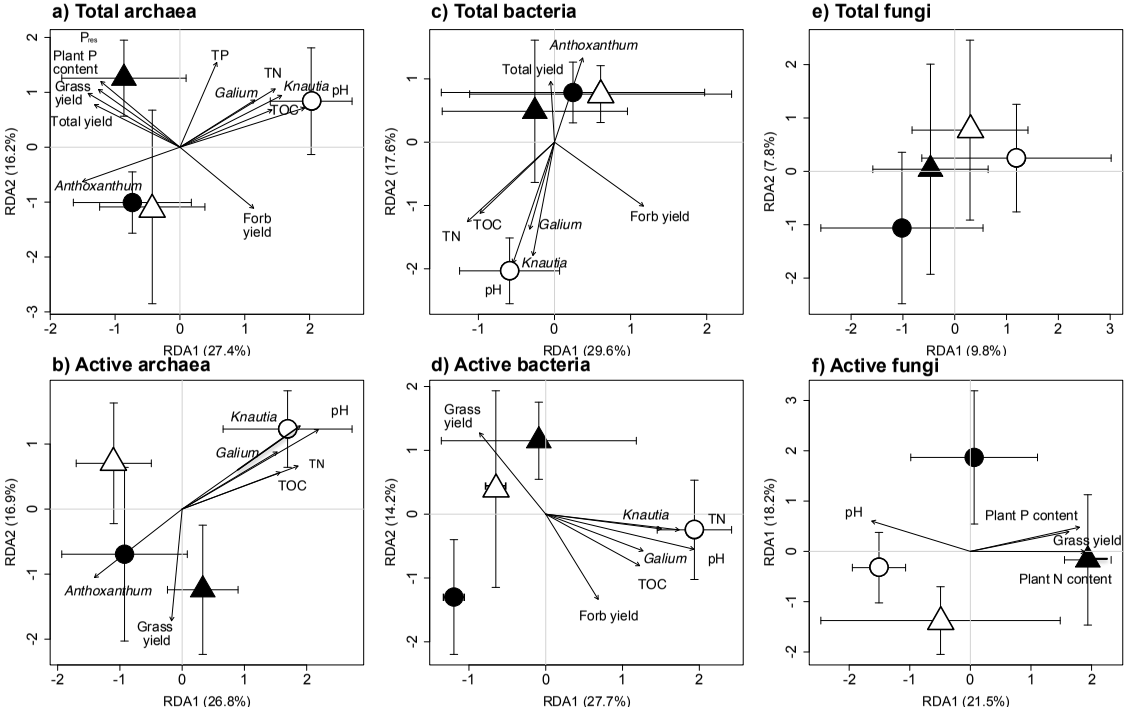


Figure 5.2. Redundancy analysis (RDA) of the community structures of archaea (a, b), bacteria (c, d) and fungi (e, f) based on the T-RF profiles in NK (●) and NPK (▲) and at pH 5.0 (△) and 6.0 (○) (means of four field replicates with standard deviation). Displayed vectors represent the soil and plant properties that are significantly correlated to the community structures (p-value<0.05).

5.3.4 Correlations between active archaeal, bacterial, fungal and *phoD*-harboring community structures and environmental factors

Based on the first and second component of the RDA, the total archaeal community structure in the NPK treatment was positively correlated to P_{res}, plant P content, grass yield and total plant yield, and negatively correlated to forb yield (Figure 5.2a and b, Table 5.4). In addition to pH, the total archaeal community structure at pH 6.1 was also positively correlated to TOC, TN and the abundances of *Galium album* and *Knautia arvensis*, while the community structure in NK and at pH 5.0 was negatively related to these variables. The abundance of *Galium album* decreased with pH and was thus negatively related to the communities at pH 6.1. The active archaeal community structure was correlated to the same plant species as the total archaeal community, and was mainly affected by pH, TOC and TN. Both total and active bacterial community structures showed a strong separation between the two pH levels and to a lesser degree between the fertilization treatments, which was mostly correlated with pH, TOC, TN and the abundances of *Galium album* and *Knautia arvensis* (Figure 5.2a and d, Table 5.4). In contrast, the total fungal community structure was not correlated to any of the soil and

Table 5.3. Analysis of similarity comparing total and active archaeal, bacterial, fungal and *phoD*-harboring communities between fertilization treatments and pH levels. Significant differences are indicated by * (p-value<0.05).

Community levels		Fertilization treatments	pH
Archaea	Total	0.03*	0.03*
	Active	0.04*	0.02*
Bacteria	Total	0.07	0.02*
	Active	0.02*	0.02*
Fungi	Total	0.40	0.46
	Active	0.03*	0.02*
phoD	Total	0.02*	0.01*
	Active	0.01*	0.01*

plant properties (Figure 5.2e, Table 5.4). The active fungal community structure, however, was related to soil pH as well as to plant N and P content and grass yield, which were all significantly higher in NPK than in NK treatment (Figure 5.2f, Table 5.4).

Table 5.4. Correlation between microbial community structures and soil and plant properties. *p*-values were retrieved from the RDA on the community structures.

	Archaea		Bacteria		Fungi		*phoD*	
	Total	Active	Total	Active	Total	Active	Total	Active
pH	0.002	0.001	0.001	0.001	n.s.	0.027	0.083	0.094
TN	0.006	0.009	0.001	0.028	n.s.	n.s.	0.097	n.s.
TOC	0.023	0.045	0.016	0.043	n.s.	n.s.	0.093	0.1
TP	0.023	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
P_{org}	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
P_{res}	0.005	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Grass yield	0.003	0.044	n.s.	0.026	n.s.	0.022	n.s.	n.s.
Legume yield	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Forb yield	0.032	n.s.	0.011	0.039	n.s.	n.s.	n.s.	n.s.
Total yield	0.023	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Plant N uptake	n.s.	n.s.	n.s.	n.s.	n.s.	0.014	n.s.	n.s.
Plant P uptake	0.005	n.s.	n.s.	n.s.	n.s.	0.046	n.s.	n.s.
Plant richness	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Anthoxanthum odoratum	0.015	0.031	0.049	n.s.	n.s.	n.s.	n.s.	n.s.
Arrhetherum elatius	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Cynosurus cristatus	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Dactylis glomerata aggr.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Galium album	0.047	0.035	0.034	0.07	n.s.	n.s.	n.s.	n.s.
Knautia arvensis	0.001	0.001	0.001	0.002	n.s.	n.s.	n.s.	n.s.
Lotus corniculatus	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Ranunculus acris	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Trifolium pratense	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

The total *phoD*-harboring community structure was correlated with pH, TOC and TN, whereas the active *phoD*-harboring community structure was only correlated with pH and TOC. In both cases, these soil properties were mainly associated with the first component, which explained 19.9 and 16.6% of the variability in the total and active *phoD*-harboring community structures, respectively (Figure 5.3 a and b, Table 5.4). Additionally, the active *phoD*-harboring community structure of the NK and NPK treatments showed tighter clustering than for the total *phoD*-harboring community structure, suggesting a more similar community structure of the active than of the total *phoD*-harboring members

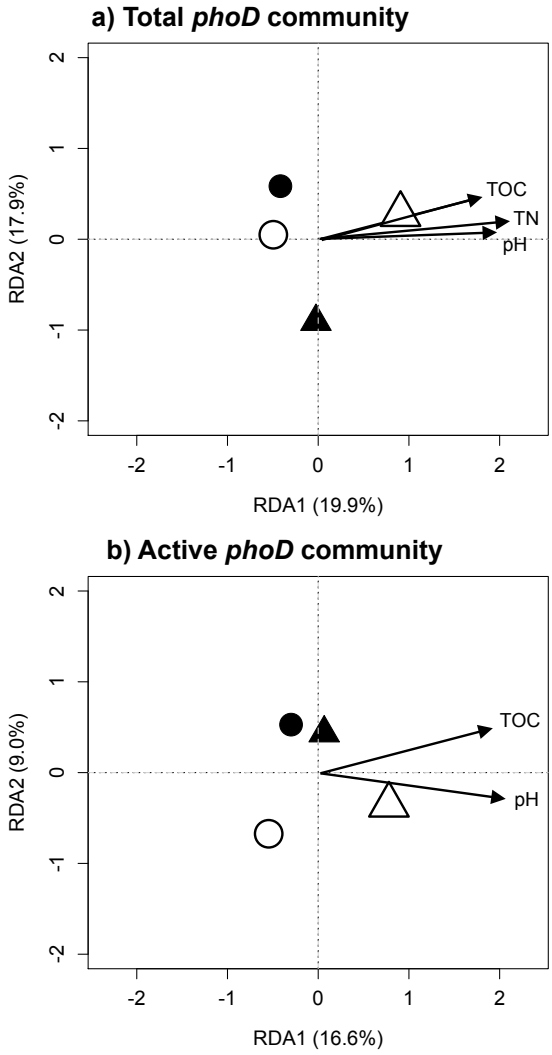


Figure 5.3. Redundancy analysis (RDA) of total and active *phoD*-harboring community structures based on 454-sequencing data in NK (●) and NPK (▲) and at pH 5.0 (△) and 6.0 (○). Displayed vectors represent the soil and plant properties that are significantly correlated to the community structures p-value<0.1).

(Figure 5.3a and b). The distribution of the total and active *phoD*-harboring community structures was similar at both pH levels (Figure 5.3a and b).

5.3.5 Taxonomic composition of the total and active *phoD*-harboring community as related to soil and plant properties

The taxonomic composition of the total and active *phoD*-harboring community was investigated using 454-sequencing and BLAST+ to elucidate the key *phoD*-harboring members in the studied soils. Taxonomy was assigned to the majority of OTUs, which were all affiliated with the bacterial kingdom. A remainder of 692 reads could not be identified and represented 0.8-3.1% of the reads before normalization (Table S5.4 and S5.5). The *phoD* gene was found in 28 bacterial orders (Figure 5.4). The dominant bacterial orders across all samples were *Actinomycetales* (15-32%), *Bacilliales* (5-15%), *Gloeobacterales* (6-23%), *Planctomycetales* (17%-30%) and *Rhizobiales* (5%-26%).

The dominant bacterial orders that harbor and express the *phoD* gene were similar in all plots. Nonetheless, some differences between treatments and/or between the total and active community were observed. For example, *Rhizobiales* had a low relative abundance in the NK treatment, while they were dominant in the NPK treatment (Figure 5.4). *Xanthomonadales* had a low relative abundance in the total community, but a high relative abundance in the active community of the NK treatment. Finally, the relative abundance of some groups such as *Gloeobacterales* was not affected by the treatments.

Furthermore, the relative abundance of several *phoD*-harboring orders was strongly correlated to individual soil properties (Figure 5.4). The abundance of *Caulobacteriales* and *Pseudomonadales* decreased with increasing P_{org} and P_{res}, while that of *Planctomycetales* increased with P_{org} and P_{res}. Additionally, the abundance of *Actinomycetales* and *Gemmantimonadales* increased together with pH, TN, TOC and TP, while the abundance of *Burkholderiales* decreased with increasing pH, TN and TP, and the abundance

of *Gloeobacterales* decreased with increasing pH, TOC and TN. All correlated soil properties were in turn affected by either phosphate depletion or pH.

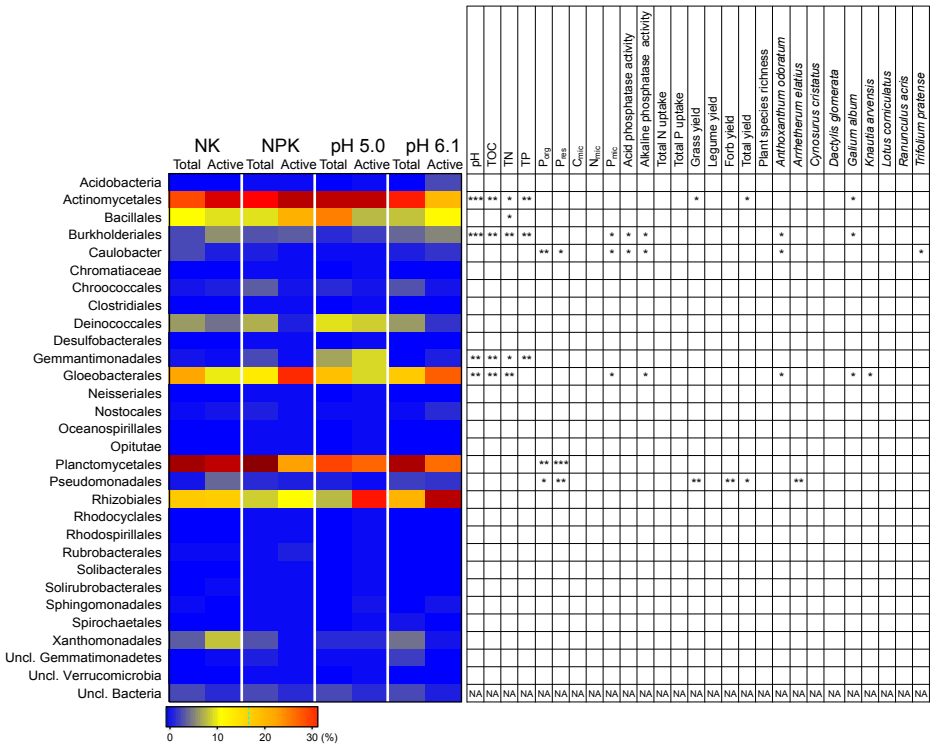


Figure 5.4. Heatmap representing the relative abundances of the total and active *phoD*-harboring orders. The table shows the significant correlations between the relative abundance of each order with soil and plant properties with *p*-values of <0.05, <0.01 and <0.001 indicated by *, ** and ***, respectively.

5.4 Discussion

5.4.1 Effect of phosphate depletion on the total and active archaeal, bacterial, fungal and *phoD*-harboring community structures

In this work, we studied the effects of phosphate depletion and soil pH on the archaeal, bacterial, fungal and *phoD*-harboring community structures in a long-term fertilization trial characterized by a pH gradient on site. Our results show that the active archaeal, bacterial and fungal community structures are more responsive to phosphate depletion than their corresponding total community structures, except for the total archaeal community structure, which was affected by phosphate depletion and correlated to phosphate availability (P_{res}). Indicators of phosphate depletion (TP, $P_{org}$and P_{res}) were otherwise not significant drivers of the communities in the first and second components of the RDA (Figure 5.2, Table 5.4). The active community structures of archaea, bacteria and fungi were all correlated with pH and grass yield, which is an indicator of productivity (Figure 5.2, Table 5.4).

Our results are in agreement with those of He et al. (2007) who showed a significant shift in total archaeal community structure between NK and NPK treatments in a wheat-maize rotation cropping system. In their study, phosphate addition specifically promoted ammonia-oxidizing *Crenarchaeota*. Our results on the total bacterial community, however, contrast with several studies that reported a significant effect of phosphate addition on the total bacterial community structure in grassland, cropped and forest soils (Zhong and Cai 2007; Beauregard et al. 2010; Liu et al. 2012; Tan et al. 2013). The total fungal community has also been reported to be shaped primarily by phosphate availability (Lauber et al. 2008). The fact that in our study only the active bacterial and fungal community responded to phosphate depletion may be related to the level of phosphate fertilization. Previous studies in the same long-term trial revealed that microorganisms were P limited in the NK treatment (Bünemann et al. 2012) and that the P balance equaled

zero in NPK treatment, suggesting that phosphate addition was adequate to maintain plant productivity but did not lead to a net increase in available and total P (Liebisch et al. 2013). Our results suggest that phosphate depletion over time impacts only the active bacteria and fungi, but does not change the structure of the greater total community. Our data also suggest that archaea are more sensitive to phosphate depletion than bacteria and fungi.

Since *phoD* is part of the Pho regulon (Vershinina and Znamenskaya 2002), we extracted available P (P_{res}) to be correlated with the active *phoD*-harboring community structure. Moreover, low available P (P_{res}) was expected to result in an increase of acid/alkaline phosphatase activity in NK compared to NPK. Our results showed that phosphate depletion led to a significant shift in the total and active *phoD*-harboring community structure and increased acid phosphatase activity (Figure 5.3, Table 5.4).However, this shift in the total and active *phoD*-harboring community structure was not correlated to any of the phosphate depletion indicators (TP, P_{org} and P_{res}). Based on the first and second RDA components, our results show that the total and active *phoD*-harboring community structures were most strongly correlated to pH and TOC, suggesting that the effect of phosphate depletion may be masked by the strong effect of pH and TOC in the RDA. Previous DNA-based studies on *phoD* have shown contrasting results of phosphate addition. Diversity of *phoD* has been reported to increase, decrease and not to be affected by phosphatase addition (Chhabra et al. 2013; Tan et al. 2013; Jorquera et al. 2014). In a long-term trial including non-fertilized soils and soil fertilized with composted manure or water-soluble phosphate, Fraser et al. (2015b) observed a general positive correlation between *phoD* gene abundance and potential alkaline phosphatase activity in soil, suggesting that *phoD* contributes significantly to the total alkaline phosphatase activity. However, comparing the non-fertilized and the water-soluble fertilized soils, they did not observe a consistent response to phosphate addition in alkaline phosphatase activity and *phoD* gene abundance over two years of the experiment, suggesting that water-soluble phosphate addition does not always decrease potential alkaline phosphatase activity. Our results are

in agreement with the outcome of their study and showed that phosphate depletion did not increase alkaline phosphatase activity nor change the total *phoD*-harboring community structure as hypothesized. However, our results suggest that phosphate depletion affects significantly the active *phoD*-harboring community composition, even though phosphate depletion indicators (TP, P_{org} and P_{res}) were not significant drivers in the first and second components of the RDA (Figure 5.3, Table 5.4).

5.4.2 Effect of soil pH on the total and active archaeal, bacterial, fungal and *phoD*-harboring community structures

One unit difference in pH value affected all investigated total and active microbial community structures except for the total fungal community (Figure 5.2, Figure 5.3, Table 5.4). Soil pH also changed some soil and plant properties such as TOC and TN, which in turn were important drivers of the microbial community structures. While bacterial community structure is widely known to be strongly correlated with soil pH (Fierer and Jackson 2006; Lauber et al. 2008; Griffiths et al. 2011; Nacke et al. 2011), less information is available for archaea and fungi. Studies of *Crenarchaeota* (Lehtovirta et al. 2009) and crenarchaeal ammonium-oxidizing archaea (Nicol et al. 2008) showed that pH can select for particular archaeal groups. Lauber et al. (2008) reported that the total fungal community structure is stable over pH values ranging from 3.6 to 4.7, though it is possible that only specific fungal taxa such as *Helotiales* and *Hypocreales* are responsive to pH changes (Rousk et al. 2010). Our results suggest that the total fungal community is also stable over pH values ranging from 5.0 to 6.0, while, one unit difference in pH value affected the active fungal community. This suggests that the active fungal community is more sensitive to pH differences than the greater total fungal community. More importantly, we could not distinguish the effect of pH from that of TOC and total N, which are both commonly correlated with pH in soil (Drenovsky et al. 2004; Lauber et al. 2008; Siciliano et al. 2014). Previous studies have reported pH as the

main driver of the total *phoD*-harboring community in arable and grassland soils (Wang et al. 2012b; Cui et al. 2015; Ragot et al. 2015). This suggests that pH rather than TOC and TN led to a shift in the structure of the total and active *phoD*-harboring community.

5.4.3 Key *phoD*-harboring microorganisms

The *phoD* gene was only found in bacteria in this study, although it is also present in archaea (e.g. *Euryarcheota*) and fungi (e.g: *Ascomycetes*) (see Integrated Microbial Genomes database, https://img.jgi.doe.gov/). Our results showed that despite differences in soil and plant properties, the dominant bacterial orders that harbor and express the *phoD* gene were similar in all plots, suggesting that the composition of the total and active *phoD*-harboring community is rather stable. Nonetheless, differences between treatments in relative abundances of both the total and active communities were also observed. For example, *phoD*-harboring *Gemmatimonadales* were dominant at pH 5.0 in both the total and active communities, whereas they were almost absent at pH 6.1 (Figure 4). Furthermore, differences between total and active *phoD*-harboring community compositions within one treatment indicate that several orders are more active than others under certain soil conditions. For example, the relative abundance of *phoD*-harboring *Xanthomonadales* was higher in the active than in the total community of the NK treatment, while it was lower in the active than in the total community of the NPK treatment. Similarly, the active *phoD*-harboring *Planctomycetes* and *Pseudomonadales* had a lower relative abundance in the NPK than in the NK treatment, while they had a similar relative abundance in the total community. Additionally, the relative abundance of the total and the active *phoD*-harboring *Bacilliales* at pH 5.0 and 6.1 suggests that higher pH strongly up-regulates the expression of the *phoD* gene in this taxonomic group. Additionally, strong correlations were observed between soil pH, TOC, TP and TN and the relative abundance of some bacterial orders such as *Actinomycetales* (Figure 4).

In contrast to our results, Tan et al. (2013) observed that water-soluble phosphate addition decreased the relative abundance of *phoD*-harboring *Acidobacteria* and *Pseudomonas fluorescens*. They reported that the dominant *phoD*-harboring phyla were the same at all studied phosphate fertilization regimes (0, 15 and 30 kg P ha^{-1}) and included mainly *Proteobacteria* (predominantly *Alphaproteobacteria*), *Actinobacteria* and *Cyanobacteria*. In our study, the key *phoD*-harboring phyla were *Actinobacteria*, *Cyanobacteria*, *Firmicutes* and *Planctomycetes*. The differences in the *phoD*-harboring community composition between our results and those of Tan et al. (2013) may be attributed to the fact that Tan et al. (2013) used primers designed by Sakurai et al. (2008), which have been shown to have an amplification bias, amplifying only a limited number of microbial groups and resulting in an over-representation of *Alphaproteobacteria* in the *phoD*-harboring community (Tan et al. 2013; Ragot et al. 2015). Nonetheless, our findings show that the regulation of *phoD* expression in *Bacilliales, Burkholderiales, Chroococcales, Gloeobacterales, Gemmatimonadales , Planctomycetales, Rhizobiales* and *Xanthomonadales* may be linked to phosphate depletion and pH.

In support of our findings, some of the key *phoD*-harboring taxonomic groups identified in our study are known to play an important role in other nutrient cycles. For example, *Proteobacteria* (e.g. *Rhizobiales*) are known to be important players in C cycling as decomposers (Štursová et al. 2012) and in N cycling as N_2-fixing bacteria (Hayatsu et al. 2008) . *Actinobacteria* (e.g. *Actinomycetales*) and *Cyanobacteria* (e.g. *Gloeobacterales*) also play a relevant role in N cycling as N_2-fixing and nitrifiers (Fierer et al. 2007; Hayatsu et al. 2008). Moreover, although their ecological role is not clear, *Deinococcus-Thermus* and *Planctomycetales* have been reported to grow well in high organic matter soils (Buckley et al. 2007; Štursová et al. 2012), suggesting that they also play a role as decomposers. Our results show that these taxonomic groups play an additional ecological function as organic P mineralizers in soil.

5.5 Conclusions

Our results showed that the active communities are generally more responsive to pH and phosphate depletion and, thus, may reflect better the response of microbial communities to their environment. Additionally, our results suggest that the fungal community is less responsive to pH and phosphate depletion than the archaeal, bacterial and *phoD*-harboring communities. Soil pH, C_{org} and TN were more important drivers than P status (TP, P_{org} and P_{res}) which did not appear in the first and second components of the RDA. Despite differences in soil and plant properties, the dominant *phoD*-harboring members were overall the same in all plots (*Actinomycetales, Rhizobiales, Gloeobacterales* and *Planctomycetales*). This suggests that the potential key player in organic phosphorus mineralization remain similar despite differences in certain soil conditions.

5.6 Acknowledgements

The authors would like to thank Dr. Stefan Zoller for the Perl scripts for taxonomy analysis and the Genetic Diversity Center (Zürich, Switzerland) for technical assistance. We also thank Monika Mascai for vegetation analyses, and Laurie Schönholzer for total C and N and TOC/N analyses. This work was supported by the Swiss National Science Foundation (SNF, project number 140900).

5.7 Supplementary material

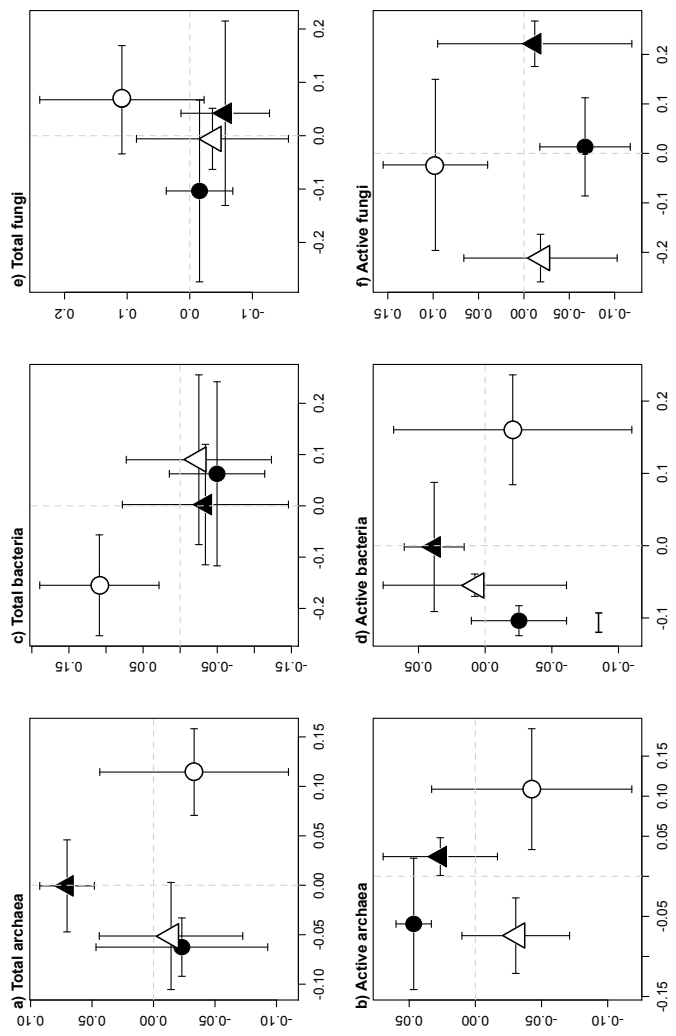


Figure S5.1. NMDS of total and active archaeal, bacterial and fungal community structures in NK (●) and NPK (▲) treatments and at pH 5.0 (△) and 6.0 (○) levels (means of four field replicates ± standard deviation).

Table S5.1. Vegetation composition, with relative abundance of each species, including results of a two tailed *t*-test comparing plant species abundance between fertilization treatments and pH levels (n.s. non-significant, *, ** and *** indicate *p*-values <0.05, <0.01, < 0.001, respectively). Species with a *p*-value <0.05 and an abundance >2% of the total community, indicated in bold, were selected for multivariate analysis.

Plant species	NK	NPK	*t*-test on phosphate depletion	pH 5.0	pH 6.1	*t*-test on pH levels
Ajuga reptans	0.5 ± 0.1	0.2 ± 0.2	n.s.	0.5 ± 0.0	0.8 ± 0.5	n.s.
Alchemilla xanthochlora	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Anthoxanthum odoratum	16.2 ± 2.5	15.3 ± 4.7	n.s.	16.7 ± 2.5	10.1 ± 2.6	*
Anthyllis vulneraria	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Arrhenatherum elatius	7.8 ± 2.5	15.1 ± 1.9	**	4.9 ± 4.9	6.3 ± 4.5	n.s.
Bellis perennis	0.2 ± 0.2	0.1 ± 0.2	n.s.	0.3 ± 0.2	0.5 ± 0.1	n.s.
Campanula patula	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Cardamine pratensis	0.1 ± 0.1	0.3 ± 0.2	n.s.	0.1 ± 0.2	0.1 ± 0.2	n.s.
Carum carvi	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Centaurea jacea	0.8 ± 0.6	0.1 ± 0.1	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Cerastium fontanum	0.2 ± 0.2	0.3 ± 0.2	n.s.	0.1 ± 0.2	0.4 ± 0.2	n.s.
Colchicum autumnale	0.4 ± 0.7	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.2 ± 0.2	n.s.
Crepis biennis	0.2 ± 0.2	0.2 ± 0.2	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Crepis capillaris	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Cynosurus cristatus	3.2 ± 1.7	2.5 ± 1.7	n.s.	8.5 ± 3.4	2.6 ± 1.0	*
Dactylis glomerata	10.9 ± 4.4	12.7 ± 2.8	n.s.	4.0 ± 0.4	6.6 ± 1.8	*
Daucus carota	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Festuca pratensis	1.1 ± 0.5	0.6 ± 0.5	n.s.	0.7 ± 0.6	2.4 ± 1.9	n.s.
Festuca rubra	0.7 ± 0.5	0.8 ± 0.6	n.s.	0.6 ± 0.8	1.1 ± 1.2	n.s.
Galium album	2.8 ± 1.1	3.5 ± 2.3	n.s.	0.9 ± 0.4	5.9 ± 2.7	*
Glechoma hederacea	0.1 ± 0.1	0.1 ± 0.1	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Helictotrichon pubescens	0.3 ± 0.3	1.5 ± 2.3	n.s.	0.1 ± 0.2	2.7 ± 3.9	n.s.
Holcus lanatus	9.5 ± 3.5	15.3 ± 2.7	**	13.8 ± 1.4	9.8 ± 4.8	n.s.
Hypochaeris radicata	1.2 ± 0.8	0.2 ± 0.2	*	2.7 ± 2.3	0.5 ± 0.1	n.s.
Knautia arvensis	0.5 ± 0.4	0.3 ± 0.2	n.s.	0.4 ± 0.2	11.3 ± 6.9	*
Lathyrus pratensis	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Leontodon hispidus	0.1 ± 0.1	0.0 ± 0.0	n.s.	0.4 ± 0.6	1.4 ± 1.5	n.s.
Leucanthemum vulgare	4.9 ± 8.4	0.2 ± 0.2	n.s.	0.0 ± 0.0	0.9 ± 0.7	*
Lolium perenne	1.1 ± 0.5	1.6 ± 0.8	n.s.	1.4 ± 0.5	1.9 ± 1.3	n.s.
Lotus corniculatus	12.9 ± 4.3	3.7 ± 0.7	**	14.8 ± 4.2	9 ± 0.6	n.s.
Medicago lupulina	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Onobrychis viciifolia	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Picris hieracioides	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Pimpinella major	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Plantago lanceolata	4.6 ± 1.7	2.9 ± 1.2	n.s.	2.7 ± 0.8	5.7 ± 2.4	n.s.
Poa trivialis	0.0 ± 0.0	1.3 ± 0.9	**	0.0 ± 0.0	0.1 ± 0.3	n.s.
Potentilla sterilis	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Prunella vulgaris	0.1 ± 0.2	0.0 ± 0.0	n.s.	0.3 ± 0.2	0.5 ± 0.1	n.s.
Ranunculus acris	2 ± 1.3	1.8 ± 0.8	n.s.	7.4 ± 1.0	1.4 ± 1.0	***

Table S5.1. Vegetation composition, with relative abundance of each species, including results of a two tailed *t*-test comparing plant species abundance between fertilization treatments and pH levels (n.s. non-significant, *, ** and *** indicate *p*-values <0.05, <0.01, < 0.001, respectively). Species with a *p*-value <0.05 and an abundance >2% of the total community, indicated in bold, were selected for multivariate analysis. (Continued)

Plant species	NK	NPK	*t*-test on phosphate depletion	pH 5.0	pH 6.1	*t*-test on pH levels
Ranunculus bulbosus	2.8 ± 1.1	1.2 ± 0.6	n.s.	2.1 ± 1.3	1.7 ± 1.6	n.s.
Rhinanthus alectorolophus	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.1 ± 0.2	n.s.
Rumex acetosa	1.9 ± 1.2	2.8 ± 1.9	n.s.	3.2 ± 0.3	0.2 ± 0.2	*
Salvia pratensis	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	1.3 ± 1.6	n.s.
Silene flos-cuculi	0.0 ± 0.0	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Taraxacum officinale	0.9 ± 0.6	0.8 ± 0.3	n.s.	1.1 ± 0.4	2 ± 0.5	n.s.
Tragopogon pratensis	0.3 ± 0.2	0.2 ± 0.2	n.s.	0.1 ± 0.2	2.2 ± 1.7	n.s.
Trifolium dubium	0.5 ± 0.6	0.0 ± 0.0	n.s.	0.6 ± 0.8	0.3 ± 0.3	n.s.
Trifolium pratense	6.5 ± 1.9	2.5 ± 0.7	*	5.9 ± 2.2	4.5 ± 0.3	n.s.
Trifolium repens	2.6 ± 1.0	2.1 ± 0.9	n.s.	4.9 ± 1.8	2.6 ± 1.3	n.s.
Trisetum flavescens	1.9 ± 1.1	8.3 ± 4.0	n.s.	1.3 ± 1.5	1.4 ± 0.9	n.s.
Veronica arvensis	0.0 ± 0.0	0.1 ± 0.1	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Veronica chamaedrys	0.2 ± 0.2	0.1 ± 0.1	n.s.	0.1 ± 0.2	0.4 ± 0.2	n.s.
Veronica serpyllifolia	0.1 ± 0.2	0.0 ± 0.0	n.s.	0.0 ± 0.0	0.0 ± 0.0	n.s.
Vicia cracca	0.0 ± 0.0	0.2 ± 0.3	n.s.	0.0 ± 0.0	0.2 ± 0.3	n.s.
Vicia sepium	0.0 ± 0.0	1.1 ± 0.6	**	0.0 ± 0.0	0.0 ± 0.0	n.s.

Table S5.2. Correlation between soil properties and selected plant abundances across the entire dataset. Upper diagonal part contains correlation coefficient estimates and lower diagonal part contains corresponding *p*-values.

	pH	TOC	TN	TP	P_{org}	P_{res}	C_{mic}	N_{mic}	P_{mic}	Acid phosphatase activity	Alkaline phosphatase activity	Total N uptake	Total P uptake	Grass yield	Legume yield	Forb yield	Total yield	Plant species richness	*Anthoxanthum odoratum*	*Arrhetherum elatius*	*Cynosurus cristatus*	*Dactylis glomerata*	*Galium album*	*Knautia arvensis*	*Lotus corniculatus*	*Ranunculus acris*	*Trifolium pratense*
pH	*****	0.890	0.841	0.565	-0.453	-0.277	0.777	0.568	0.773	-0.660	0.930	-0.396	-0.337	-0.413	0.020	0.067	-0.469	0.412	-0.649	-0.247	-0.244	0.003	0.669	0.688	-0.125	-0.513	-0.116
TOC	**<0.001**	*****	0.959	0.711	-0.262	-0.051	0.802	0.609	0.772	-0.629	0.883	-0.318	-0.185	-0.178	-0.034	-0.238	-0.311	0.607	-0.579	-0.023	-0.401	0.189	0.544	0.471	-0.238	-0.660	-0.152
TN	**<0.001**	**<0.001**	*****	0.613	-0.256	-0.106	0.733	0.629	0.674	-0.446	0.836	-0.400	-0.266	-0.238	-0.136	-0.171	-0.367	0.653	-0.560	0.000	-0.428	0.246	0.556	0.409	-0.171	-0.741	-0.116
TP	0.022	0.002	0.012	*****	0.261	0.455	0.597	0.400	0.717	-0.597	0.609	0.163	0.334	0.271	-0.153	-0.274	0.203	0.572	-0.336	0.191	-0.497	0.212	0.447	0.259	-0.600	-0.497	-0.599
Porg	0.078	0.326	0.338	0.328	*****	0.798	-0.416	-0.235	-0.214	0.056	-0.409	0.584	0.755	0.800	-0.331	-0.363	0.786	-0.050	0.241	0.518	-0.128	0.511	-0.218	-0.449	-0.520	-0.028	-0.229
Pres	0.300	0.852	0.695	0.077	**<0.001**	*****	-0.111	0.019	0.095	-0.123	-0.268	0.769	0.912	0.927	-0.162	-0.611	0.858	0.033	0.088	0.488	-0.280	0.510	-0.087	-0.276	-0.605	-0.148	-0.384
MicC	**<0.001**	**<0.001**	0.001	0.015	0.109	0.683	*****	0.810	0.832	-0.576	0.860	-0.169	-0.134	-0.218	-0.102	0.046	-0.255	0.577	-0.632	-0.005	-0.242	-0.258	0.564	0.400	-0.065	-0.444	-0.302
MicN	0.022	0.012	0.009	0.125	0.381	0.944	**<0.001**	*****	0.673	-0.407	0.682	-0.118	-0.070	-0.089	-0.262	-0.035	-0.147	0.418	-0.466	0.017	-0.041	-0.195	0.306	0.222	0.019	-0.375	-0.291
MicP	**<0.001**	**<0.001**	0.004	0.002	0.426	0.726	**<0.001**	0.004	*****	-0.617	0.730	0.032	0.011	-0.032	-0.112	-0.030	-0.062	0.522	-0.612	0.207	-0.324	-0.063	0.508	0.504	-0.276	-0.486	-0.496
Acid phosphatase activity	0.005	0.009	0.083	0.015	0.836	0.650	0.020	0.117	0.011	*****	-0.671	0.063	-0.075	-0.003	-0.231	0.260	0.076	-0.041	0.528	0.207	-0.053	0.072	-0.377	-0.505	0.215	0.095	0.020
Alkaline phosphatase activity	**<0.001**	**<0.001**	**<0.001**	0.012	0.116	0.316	**<0.001**	0.004	0.001	0.004	*****	-0.406	-0.333	-0.413	-0.108	0.130	-0.456	0.496	-0.649	-0.204	-0.261	-0.188	0.647	0.620	-0.159	-0.470	-0.211
Total N uptake	0.129	0.230	0.125	0.546	0.017	**<0.001**	0.531	0.664	0.906	0.815	0.119	*****	0.908	0.871	-0.104	-0.221	0.949	0.071	0.242	0.596	-0.325	0.248	-0.047	-0.255	-0.637	0.080	-0.444
Total P uptake	0.202	0.492	0.320	0.206	0.001	**<0.001**	0.620	0.796	0.969	0.784	0.208	**<0.001**	*****	0.957	-0.149	-0.443	0.961	0.070	0.215	0.472	-0.286	0.433	-0.034	-0.306	-0.621	-0.060	-0.355
Grass yield	0.111	0.510	0.375	0.309	**<0.001**	**<0.001**	0.417	0.743	0.905	0.992	0.112	**<0.001**	**<0.001**	*****	-0.125	-0.583	0.960	0.073	0.274	0.577	-0.266	0.533	-0.189	-0.417	-0.603	-0.066	-0.276
Legume yield	0.942	0.902	0.616	0.572	0.211	0.549	0.708	0.327	0.680	0.389	0.691	0.701	0.582	0.646	*****	-0.366	-0.194	-0.361	0.141	-0.422	0.432	0.010	-0.142	-0.134	0.213	0.568	0.528
Forb yield	0.805	0.374	0.526	0.304	0.167	0.012	0.867	0.898	0.911	0.332	0.631	0.410	0.086	0.018	0.163	*****	-0.344	0.075	-0.131	-0.061	-0.033	-0.637	0.299	0.320	0.183	0.087	-0.230
Total yield	0.067	0.240	0.161	0.451	**<0.001**	**<0.001**	0.341	0.587	0.819	0.779	0.076	**<0.001**	**<0.001**	**<0.001**	0.470	0.193	*****	0.081	0.292	0.628	-0.289	0.391	-0.124	-0.389	-0.632	0.012	-0.370
Plant species richness	0.112	0.013	0.006	0.021	0.854	0.903	0.019	0.107	0.038	0.880	0.051	0.794	0.798	0.789	0.170	0.781	0.765	*****	-0.150	0.413	-0.751	0.121	0.484	0.110	-0.432	-0.693	-0.502
Anthoxanthum odoratum	0.007	0.019	0.024	0.203	0.368	0.745	0.009	0.069	0.012	0.035	0.007	0.367	0.423	0.305	0.603	0.630	0.273	0.579	*****	-0.081	0.127	0.182	-0.711	-0.595	0.029	0.395	-0.002
Arrhetherum elatius	0.356	0.933	0.999	0.478	0.040	0.055	0.985	0.951	0.442	0.442	0.448	0.015	0.065	0.019	0.104	0.823	0.009	0.112	0.765	*****	-0.491	0.254	-0.006	-0.284	-0.510	-0.255	-0.425
Cynosurus cristatus	0.362	0.124	0.099	0.050	0.638	0.293	0.366	0.880	0.221	0.845	0.329	0.219	0.282	0.320	0.095	0.904	0.277	0.001	0.640	0.053	*****	-0.311	-0.487	-0.307	0.700	0.720	0.577
Dactylis glomerata aggr.	0.990	0.482	0.359	0.430	0.043	0.044	0.335	0.468	0.818	0.791	0.485	0.355	0.094	0.034	0.971	0.008	0.135	0.655	0.499	0.343	0.241	*****	-0.061	-0.253	-0.351	-0.419	0.138
Galium album	0.005	0.029	0.025	0.083	0.418	0.748	0.023	0.249	0.045	0.150	0.007	0.863	0.901	0.482	0.599	0.261	0.647	0.057	0.002	0.983	0.056	0.823	*****	0.636	-0.343	-0.527	-0.181
Knautia arvensis	0.003	0.066	0.116	0.333	0.081	0.301	0.125	0.409	0.046	0.046	0.010	0.340	0.249	0.108	0.621	0.227	0.136	0.686	0.015	0.286	0.247	0.344	0.008	*****	-0.149	-0.400	-0.189
Lotus corniculatus	0.644	0.374	0.527	0.014	0.039	0.013	0.812	0.945	0.301	0.425	0.556	0.008	0.010	0.013	0.427	0.499	0.009	0.095	0.914	0.044	0.003	0.182	0.194	0.582	*****	0.295	0.578
Ranunculus acris	0.042	0.005	0.001	0.050	0.917	0.585	0.085	0.152	0.056	0.726	0.066	0.770	0.826	0.809	0.022	0.747	0.964	0.003	0.130	0.341	0.002	0.106	0.036	0.125	0.267	*****	0.311
Trifolium pratense	0.668	0.573	0.670	0.014	0.393	0.142	0.256	0.275	0.051	0.941	0.433	0.085	0.177	0.300	0.036	0.391	0.158	0.047	0.995	0.100	0.019	0.611	0.502	0.482	0.019	0.241	*****

Table S5.3. Species richness of archaeal, bacterial and fungal communities based on T-RF profiles and results of *t*-test comparing species richness between total and active communities with levels of significance indicated by *, **, *** and n.s. for *p*-values <0.05, <0.01, <0.001 and non-significant, respectively. Superscript letters indicate significant differences within fertilization treatments and pH levels.

Plots	Archaea			Bacteria			Fungi		
	Total	Active	*t*-test	Total	Active	*t*-test	Total	Active	*t*-test
NK	277 ± 7	264 ± 4^{a}	**	234 ± 10	217 ± 4^{a}	***	256 ± 8^{a}	234 ± 6^{a}	***
NPK	269 ± 11	245 ± 2^{b}	**	226 ± 11	202 ± 5^{b}	***	226 ± 5^{b}	216 ± 7^{b}	n.s.
pH 5.0	278 ± 7	246 ± 6	***	230 ± 12	218 ± 5	**	230 ± 5	210 ± 8	***
pH 6.1	272 ± 13	243 ± 4	**	232 ± 10	210 ± 8	***	224 ± 9	210 ± 11	***

Table S5.4. Number of filtered reads, number of reads after normalization and species richness index (Chao1) of the total and active *phoD*- harboring community with 95% confidence intervals in fertilization treatments and pH levels.

Plot	Type of community	Number of filtered reads	Number of reads after normalization	Species richness
NK	Total	6897	999	478 ± 23
	Active	9403	1159	495 ± 26
NPK	Total	9632	1065	828 ± 36
	Active	6921	1057	273 ± 12
pH 5.0	Total	9945	1032	854 ± 37
	Active	8577	1029	353 ± 18
pH 6.1	Total	9902	1067	626 ± 30
	Active	9083	1096	152 ± 7

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels.

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
1. Acidobacteria	5	2	4	0	6	70	0	0
1.1. Acidobacteria	5	0	1	0	1	70	0	0
1.1.1. Acidobacteria	5	0	1	0	1	70	0	0
1.1.1.1. Acidobacteriaceae	5	0	1	0	1	70	0	0
1.1.1.1.1. Acidobacterium	0	0	0	0	1	66	0	0
1.1.1.1.2. Granulicella	5	0	1	0	0	4	0	0
1.2. Solibacteres	0	2	3	0	5	0	0	0
1.2.1. Solibacterales	0	2	3	0	5	0	0	0
1.2.1.1. Solibacteraceae	0	2	3	0	5	0	0	0
1.2.1.1.1.. Solibacter	0	2	3	0	5	0	0	0
1.3. uncl_Acidobacteria	0	0	0	0	0	0	0	0
2. Actinobacteria	1357	519	1518	248	477	443	285	1134
2.1. Actinomycetales	1340	510	1516	243	477	440	285	1132
2.1.1. Actinoplanaceae	605	248	812	216	213	211	133	498
2.1.1.1. Actinoplanes	89	30	77	14	11	24	11	46
2.1.1.2. Kitasatospora	0	0	2	0	0	0	0	2
2.1.1.3. Micromonospora	78	12	60	94	33	36	13	26
2.1.1.4. Microstreptospora	1	58	112	14	21	44	16	60
2.1.1.5. Streptomyces	437	144	535	94	147	94	93	353
2.1.1.6. Streptosporangium	0	4	26	0	1	13	0	11
2.1.2. Actinosynnemataceae	131	17	133	0	8	0	28	244
2.1.2.1. Actinomyces	101	2	76	0	3	0	22	198
2.1.2.2. Amycolatopsis	0	9	19	0	0	0	4	25
2.1.2.3. Kutzneria	0	0	0	0	0	0	0	0
2.1.2.4. Saccharothrix	30	6	37	0	5	0	2	21
2.1.2.5. Thermobispora	0	0	1	0	0	0	0	0
2.1.3. Beutenbergiaceae	126	50	42	0	52	28	48	6
2.1.3.1. Beutenbergia	126	50	42	0	52	28	48	6
2.1.4. Cellulomonadaceae	1	0	0	0	0	0	0	0
2.1.4.1. Cellulomonas	1	0	0	0	0	0	0	0
2.1.5. Corynebacteriaceae	0	0	0	0	1	0	0	0
2.1.5.1. Corynebacterium	0	0	0	0	1	0	0	0
2.1.6. Frankiaceae	71	44	56	2	28	34	0	23
2.1.6.1. Frankia	71	44	56	2	28	34	0	23
2.1.7. Geodermatophilaceae	0	0	1	0	0	0	0	0
2.1.7.1. Blastococcus	0	0	1	0	0	0	0	0
2.1.7.2. Geodermatophilus	0	0	0	0	0	0	0	0
2.1.7.3. Modestibacter	0	0	0	0	0	0	0	0
2.1.8. Glycomycetaceae	36	2	2	0	1	0	1	0

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels. (Continued)

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
2.1.8.1. Stackebrandtia	36	2	2	0	1	0	1	0
2.1.9. Gordoniaceae	0	0	0	0	0	0	0	2
2.1.9.1. Gordana	0	0	0	0	0	0	0	2
2.1.10. Kineosporiaceae	36	5	38	0	53	32	11	8
2.1.10.1. Kineococcus	36	5	38	0	53	32	11	8
2.1.11. Micrococcaceae	15	2	38	0	13	2	2	21
2.1.11.1. Arthrobacter	15	0	14	0	9	1	0	16
2.1.11.2. Micrococcus	0	2	24	0	4	1	2	5
2.1.12. Mycobacteriaceae	17	5	32	0	0	3	1	45
2.1.12.1. Mycobacterium	17	5	32	0	0	3	1	45
2.1.13. Nocardiaceae	4	3	4	0	1	0	0	2
2.1.13.1. Micropolyspora	0	0	0	0	0	0	0	0
2.1.13.2. Rhodococcus	4	3	4	0	1	0	0	2
2.1.14. Nocardioidaceae	245	130	350	25	99	127	61	283
2.1.14.1. Kribella	243	130	343	25	98	126	61	274
2.1.14.2. Nocardia	2	0	7	0	1	1	0	9
2.1.15. Propionibacteriaceae	0	0	0	0	0	0	0	0
2.1.15.1. Microlunatus	0	0	0	0	0	0	0	0
2.1.16. Thermomonosporaceae	53	4	8	0	8	2	0	0
2.1.16.1. Thermomonospora	53	4	8	0	8	2	0	0
2.1.17. Tsukamurellaceae	0	0	0	0	0	1	0	0
2.1.17.1. Tsukamurella	0	0	0	0	0	1	0	0
2.2. Rubrobacterales	17	4	2	5	0	3	0	2
2.2.1. Rubrobacteraceae	17	4	2	5	0	3	0	2
2.2.1.1. Rubrobacter	17	4	2	5	0	3	0	2
2.3. Solirubrobacterales	0	5	0	0	0	0	0	0
2.3.1. Conexibacteraceae	0	5	0	0	0	0	0	0
2.3.1.1. Conexibacter	0	5	0	0	0	0	0	0
3. Bacteroidetes	0	0	0	0	0	4	0	0
3.1. Sphingobacteriia	0	0	0	0	0	4	0	0
3.1.1. Chitinophagaceae	0	0	0	0	0	4	0	0
3.1.1.1. Niastella	0	0	0	0	0	4	0	0
4. Chloroflexi	0	0	0	0	0	0	0	0
4.1. Chloroflexi	0	0	0	0	0	0	0	0
4.1.1. Chloroflexaceae	0	0	0	0	0	0	0	0
4.1.1.1. Roseiflexaceae	0	0	0	0	0	0	0	0
4.1.1.1.1. Roseiflexus	0	0	0	0	0	0	0	0
4.1.2. Herpetosiphonales	0	0	0	0	0	0	0	0

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels. (Continued)

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
4.1.2.1. Herpetosiphonaceae	0	0	0	0	0	0	0	0
4.1.2.1.1. Herpetosiphon	0	0	0	0	0	0	0	0
4.2. Thermomicrobia	0	0	0	0	0	0	0	0
4.2.1. Sphaerobacterales	0	0	0	0	0	0	0	0
4.2.1.1. Sphaerobacteraceae	0	0	0	0	0	0	0	0
4.2.1.1.1. Sphaerobacter	0	0	0	0	0	0	0	0
5. Cyanobacteria	1105	173	828	188	331	673	72	610
5.1. Cyanophyceae	15	12	49	0	7	43	0	23
5.1.1. Chroococcales	0	0	0	0	0	0	0	6
5.1.1.1. Xenococcaceae	0	0	0	0	0	0	0	6
5.1.1.1. Chroococcidiopsis	0	0	0	0	0	0	0	6
5.1.1.2. Pleurocapsa	0	0	0	0	0	0	0	0
5.1.2. Nostocales	15	12	49	0	7	43	0	17
5.1.2.1. Nostocaceae	15	12	49	0	7	43	0	17
5.1.2.1.1. Amorphonostoc	0	0	2	0	0	0	0	9
5.1.2.1.2. Anabaena	15	12	47	0	7	43	0	8
5.2. Gloeobacteria	1045	141	600	186	269	605	67	542
5.2.1. Gloeobacterales	1045	141	600	186	269	605	67	542
5.2.1.1. Gloeobacter	1045	141	600	186	269	605	67	542
5.2.1.1.1. Gloeobacter	1045	141	600	186	269	605	67	542
5.3. Oscillatoriophycideae	45	20	179	2	55	25	5	45
5.3.1. Chroococcales	45	20	179	2	55	25	5	45
5.3.1.1. Chroococcales	45	20	179	2	55	25	5	45
5.3.1.1.1. Chroococcus	45	20	179	2	55	25	5	45
5.3.1.1.2. Cyanothece	0	0	0	0	0	0	0	0
5.3.2. Oscillatoriales	0	0	0	0	0	0	0	0
5.3.2.1. Oscillatoriales	0	0	0	0	0	0	0	0
5.3.2.1.1. Microcoleus	0	0	0	0	0	0	0	0
6. Deinococcus-Thermus	326	70	347	4	105	47	64	287
6.1. Hadobacteria	326	70	347	4	105	47	64	287
6.1.1. Deinococcales	326	70	347	4	105	47	64	287
6.1.1.1. Deinococcaceae	326	70	345	4	105	47	64	287
6.1.1.1.1. Deinobacter	326	70	345	4	105	47	64	287
6.1.1.2. Trueperaceae	0	0	2	0	0	0	0	0
6.1.1.2.1. Truepera	0	0	2	0	0	0	0	0
7. Firmicutes	546	138	441	124	132	263	58	711
7.1. Bacilli	538	138	441	124	132	263	58	711
7.1.1. Bacillales	538	138	441	124	132	263	58	711

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels. (Continued)

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
7.1.1.1. Bacillaceae	538	138	441	124	132	263	58	711
7.1.1.1.1. Bacillus	536	138	440	124	132	263	58	709
7.1.1.1.2. Geobacillus	2	0	1	0	0	0	0	2
7.2. Clostridia	8	0	0	0	0	0	0	0
7.2.1. Clostridiales	8	0	0	0	0	0	0	0
7.2.1.1. Peptococcaceae	8	0	0	0	0	0	0	0
7.2.1.1.1. Desulfitobacterium	8	0	0	0	0	0	0	0
8. Gemmatimonadetes	47	14	168	0	47	37	67	233
8.1. Gemmatimonadetes	47	6	126	0	5	34	67	215
8.1.1. Gemmantimonadales	47	6	126	0	5	34	67	215
8.1.1.1. Gemmantimonadaceae	0	0	126	0	5	34	67	215
8.1.1.1.1. Gemmatimonas	70	6	126	0	5	34	67	215
8.2. Unclassified Gemmatimonadetes	0	8	42	0	42	3	0	18
9. Nitrospirae	0	0	0	0	0	0	0	0
9.1. Nitrospira	0	0	0	0	0	0	0	0
9.1.1. Nitrospirales	0	0	0	0	0	0	0	0
9.1.1.1. Nitrospiraceae	0	0	0	0	0	0	0	0
9.1.1.1.1. Nitrospira	0	0	0	0	0	0	0	0
10. Planctomycetes	1982	545	2033	137	633	582	190	824
10.1. Planctomycetacia	1982	545	2033	137	633	582	190	824
10.1.1. Planctomycetales	1982	545	2033	137	633	582	190	824
10.1.1.1. Planctomycetaceae	1812	545	2033	137	633	582	190	824
10.1.1.1.1. Isosphaera	0	33	203	0	44	104	9	89
10.1.1.1.2. Pirella	171	45	192	32	46	25	13	148
10.1.1.1.3. Planctomyces	0	0	1	0	0	0	0	13
10.1.1.1.4. Rhodopirellula	20	6	16	0	7	0	0	5
10.1.1.1.5. Singulisphaera	1621	461	1621	105	536	453	168	569
11. Proteobacteria	1391	522	867	92	523	1164	269	416
11.1. Alphaproteobacteria	989	253	469	65	332	960	236	255
11.1.1. Caulobacter	146	18	51	0	23	52	0	14
11.1.1.1. Caulobacteraceae	146	18	51	0	23	52	0	14
11.1.1.1.1. Asticcacaulis	4	0	0	0	0	0	0	0
11.1.1.1.2. Brevundimonas	0	0	8	0	0	0	0	0
11.1.1.1.3. Caulobacter	2	1	2	0	0	0	0	11
11.1.1.1.4. Phenylobacterium	140	17	41	0	23	52	0	3
11.1.2. Rhizobiales	826	234	398	65	308	891	232	238
11.1.2.1. Beijerinckiaceae	0	0	0	0	0	0	5	0

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels. (Continued)

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
11.1.2.1.1. Beijerinckia	0	0	0	0	0	0	5	0
11.1.2.2. Bradyrhizobiaceae	214	56	114	12	95	105	42	37
11.1.2.2.1. Bradyrhizobium	186	48	111	7	95	105	27	34
11.1.2.2.2. Oligotropha	0	0	0	5	0	0	0	0
11.1.2.2.3. Rhodopseudomonas	28	8	3	0	0	0	3	3
11.1.2.3. Methylobacteriaceae	514	134	197	38	197	466	59	142
11.1.2.3.1. Methylobacterium	514	134	197	38	197	466	59	142
11.1.2.4. Methylocystaceae	0	0	0	0	0	0	23	0
11.1.2.4.1. Methylocystis	0	0	0	0	0	0	23	0
11.1.2.5. Phyllobacteriaceae	0	0	0	0	0	209	55	5
11.1.2.5.1. Mesorhizobium	0	0	0	0	0	209	55	5
11.1.2.6. Rhizobiaceae	9	14	10	0	5	87	34	8
11.1.2.6.1. Agrobacterium	0	0	3	0	0	23	14	3
11.1.2.6.2. Pelagibacterium	0	0	0	0	0	3	0	3
11.1.2.6.3. Rhizobium	9	14	7	0	5	55	13	0
11.1.2.6.4. Sinorhizobium	0	0	0	0	0	6	7	2
11.1.2.7. Xanthobacteraceae	0	0	0	1	0	0	3	0
11.1.2.7.1. Azorhizobium	0	0	0	0	0	0	2	0
11.1.2.7.2. Starkeya	0	0	0	1	0	0	1	0
11.1.3. Rhodobacterales	0	0	0	0	0	0	0	0
11.1.3.1. Rhodobacteraceae	0	0	0	0	0	0	0	0
11.1.3.1.1. Ketogulonicigenium	0	0	0	0	0	0	0	0
11.1.4. Rhodospirillales	3	0	6	0	0	0	0	3
11.1.4.1. Gluconobacteraceae	3	0	5	0	0	0	0	1
11.1.4.1.1. Acidiphilium	0	0	1	0	0	0	0	0
11.1.4.1.2. Gluconobacter	3	0	4	0	0	0	0	1
11.1.4.2. Rhodospirillaceae	0	0	1	0	0	0	0	2
11.1.4.2.1. Azospirillum	0	0	0	0	0	0	0	0
11.1.4.2.2. Rhodospirillum	2	2	2	2	2	2	2	2
11.1.5. Sphingomonadales	14	1	14	0	1	17	4	0
11.1.5.1. Sphingomonadaceae	14	1	14	0	1	17	4	0
11.1.5.1.1. Sphingobium	4	0	5	0	0	17	0	0
11.1.5.1.2. Sphingomonas	10	1	9	0	1	0	4	0
11.1.5.1.3. Sphingopyxis	0	0	0	0	0	0	0	0
11.2. Betaproteobacteria	157	88	157	21	70	132	18	55
11.2.1. Burkholderiales	157	88	157	21	70	132	18	55
11.2.1.1. Alcaligenaceae	2	3	18	4	0	0	0	0
11.2.1.1.1. Achromobacter	0	2	13	0	0	0	0	0
11.2.1.1.2. Bordetella	2	1	5	4	0	0	0	0

Table S5.5. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the total and active *phoD*-harboring communities of fertilization treatments and pH levels. (Continued)

	Relative abundance (nb of reads)							
	NK		NPK		pH 5.0		pH 6.1	
	Total	Active	Total	Active	Total	Active	Total	Active
11.2.1.2. Burkholderiaceae	18	21	48	1	31	24	1	30
11.2.1.2.1. Burkholderia	0	12	6	0	1	2	0	4
11.2.1.2.2. Cupravidus	0	2	13	0	0	0	1	4
11.2.1.2.3. Ralstonia	18	7	29	1	30	22	0	22
11.2.1.3. Comamonadaceae	122	34	54	16	22	106	1	18
11.2.1.3.1. Acidivorax	33	1	3	0	0	0	0	4
11.2.1.3.2. Albidoferax	11	0	4	0	2	10	1	0

6

GENERAL DISCUSSION

In this thesis, the current knowledge on the environmental prevalence and taxonomic distribution of the alkaline phosphatase genes *phoD* and *phoX* was assessed and new primer sets that amplify *phoD* and *phoX* genes present in soil microorganisms were designed (Chapters 2 and 3). Using the newly-designed primers, key microorganisms that harbor the *phoD* and *phoX* genes were identified and the relationships between the environmental factors and the *phoD*- and *phoX*-harboring community structure and composition were explored in thrity soils across three land-uses, three climate zones and six soil groups in Australia and Switzerland (Chapter 4). Finally, the effects of phosphate depletion and pH on both the total and active *phoD*-harboring community structure and composition were assessed in a long-term P fertilization trial on a permanent grassland characterized by a natural pH gradient on the site (Chapter 5). The main findings are discussed and synthesized below. First, the coverage of the primers is evaluated (6.1). The environmental prevalence (6.2) and the taxonomic distribution of *phoD* and *phoX* as well as the active and total *phoD*-harboring microorganisms (6.3) are then compared. The correlations between environmental factors and the *phoD*- and *phoX*-harboring community structure and composition observed in our soil samples are summarized (6.4). Finally, our findings are discussed in an environmental (6.5) and agronomic perspective (6.6).

6.1 Evaluation of the coverage of the primers

Primers are essential tools for environmental molecular ecology. Their design is based on our knowledge of the gene of interest. For this reason, they are unlikely to cover the entire existing diversity of the gene, as the majority of gene sequences and organisms is still not fully described. Nevertheless, primers represent one of the best tools to study the uncultivable below-ground microbial diversity in the environment due to their high sensitivity and broad applicability. The first primer set targeting *phoD* in soil microorganisms was developed by Sakurai et al. (2008) based on *phoD* phosphatase gene

sequences from seven isolates. Using these primers combined with 454-sequencing, Tan et al. (2013) and Fraser et al. (2015a) showed that these primers have an amplification bias, resulting in an overrepresentation of *Alphaproteobacteria*, and that new primers were therefore required to provide better coverage of the *phoD* diversity. To our knowledge, no primer set that targets *phoX* in soil microorganisms had been developed before.

The coverage of the *phoD*- and *phoX*-targeting primers developed in this thesis can be evaluated by comparing the diversity of the genes amplified using our primers with the diversity of *phoD* and *phoX* sequences available in databases such as the Integrative Microbial Genomes and Metagenomes (IMG/M) database. The IMG/M database is a dedicated system for annotation of whole genomes and metagenomes and represents our current knowledge of the taxonomic distribution of *phoD* and *phoX* across prokaryotes and microbial eukaryotes (Chapters 2 and 3). In the IMG/M database, the *phoD* gene was found in archaea, bacteria and fungi, while *phoX* was reported in bacteria only (Table 6.1). Using our primers, *phoD* was amplified from archaea, bacteria and fungi but from fewer bacterial and fungal phyla than reported in the IMG/M database (Table 6.1). The *phoX* gene was amplified from archaea and bacteria.

Table 6.1. Number of phyla harboring homologues of the *phoD* and *phoX* genes either based on the amplification using the newly-designed primers or reported by the IMG/M database.

Kingdom	*phoD*		*phoX*	
	Primer-amplified	IMG/M database	Primer-amplified	IMG/M database
Archaea	1	1	1	0
Bacteria	13	20	16	15
Fungi	2	3	0	0

The comparison of the number of phyla between the *phoD* and *phoX* sequences available in the IMG/M database and the results obtained by amplification with our primers shows that our newly-designed primers cover a large fraction of the *phoD* and *phoX* diversity. However, the number of phyla found using either method shows

some differences that can be due to biases created by the database, the primers or the type of studied samples. The biases created by the database originate from the fact that databases of functional genes consist of annotated genes from whole genomes, which come from cultivated microbial strains only. Secondly, the primers developed in this thesis were designed based on soil microbial sequences retrieved from databases, which consist of a limited number of available and well-annotated gene sequences. This creates automatically a bias in the coverage towards the currently known genes in soil microorganisms and may have a strong impact on the primer coverage. Finally, the primers designed in this study were only used on soil samples, while the IMG/M database includes genomes from diverse ecosystems such as aquatic and host-associated environments that may include an additional diversity of microorganisms harboring *phoD* and *phoX*, which is not found in soil.

Nonetheless, our results show that the newly-designed primers cover a large fraction of the known *phoD*- and *phoX*-harboring phyla, and thus, represent valuable tools to study phosphatase-harboring microorganisms in soil. Additionally, our results show that the *phoD*-targeting primers developed in this thesis have a better coverage of the *phoD* diversity in soil than the *phoD*-targeting primers published by Sakurai et al. (2008) (Chapter 2). Moreover, our primers were designed to target a fragment size long enough for sequencing and profiling techniques (e.g. T-RFLP), and short enough for quantitative real-time PCR (qPCR). Preliminary work showed successful amplification of *phoD* genes and transcripts from soil DNA using our primers (data not shown).

6.2 Prevalence of *phoD* and *phoX* phosphatase genes in the environment - a meta-analysis

Our current knowledge of the prevalence of the *phoD* and *phoX* genes in the environment was assessed using the metagenomes available from the IMG/M database and our results of the sequencing analysis using our primers on the soils studied in this thesis (Chapters

2, 3 and 4). The metagenomes in the database comprised 21 different environment types, including 10 environments for free-living microorganisms and 11 environments of microorganisms associated with hosts (Table 6.2).

Table 6.2. Presence of the *phoD* and *phoX* genes in different types of environments (based on the IMG/M database).

	phoD	*phoX*
Free-living		
Air	✓	✓
Aquatic	✓	✓
Bioreactor	✓	✓
Bioremediation	✓	✓
Biotransformation	✓	✓
Lab enrichment	✓	✓
Food	-	-
Solid waste	✓	✓
Terrestrial	✓	✓
Wastewater	✓	✓
Host-associated		
Algae	-	-
Annelida	✓	✓
Arthropoda	✓	✓
Birds	✓	✓
Cnideria	-	-
Human	✓	✓
Mammals	✓	✓
Mollusca	✓	✓
Plants	✓	✓
Porifera	✓	✓
Tunicates	✓	✓

The *phoD* and *phoX* genes were present in very diverse types of environments, showing that *phoD*- and *phoX*-harboring microorganisms are widely spread in the environment. They were reported from the same types of environments and were most abundant in terrestrial ecosystems. *phoD* was also found in high frequency in marine and air ecosystems and *phoX* in microbiomes associated with plants. In our sequencing studies, *phoD* was amplified in 36 of the 38 soil samples (Chapters 2, 4 and 5), while *phoX*

was amplified in 27 of the 30 soil samples studied (Chapters 3 and 4). The absence of amplification in some soils may be attributed to the absence or too low abundance of the genes to allow amplification. However, DNA quality was high enough for amplification as at least one of the studied genes was successfully amplified in all samples.

The presence of *phoD* has been reported in grassland and arable soils using the primers designed by Sakurai et al. (2008) (Chhabra et al. 2013; Fraser et al. 2015b; Sakurai et al. 2008; Tan et al. 2013; Wang et al. 2012a; Wang et al. 2012b) and in forest soils in a shot-gun metagenomic study (Bergkemper et al. 2015). However, to our knowledge, this is the first study on *phoX* in soil. The presence of *phoD* and *phoX* in marine and fresh water ecosystems has already been reported by several studies (Dai et al. 2014; Luo et al. 2009; Sebastián and Ammerman 2009). In aquatic ecosystems, alkaline phosphatase may play an even bigger role than in terrestrial ecosystems as pH is generally more alkaline than in soil (Dickson 1993). The results of the meta-analysis and our sequencing analysis on soil samples suggest that *phoD* and *phoX* are widely spread but not ubiquitous in the environment.

6.3 Key phosphatase gene-harboring microorganisms

6.3.1 Taxonomic distribution of *phoD* and *phoX*

The taxonomic distribution of the phosphatase genes was investigated by sequencing analysis using our newly-designed primers with successful amplification on 36 and 27 soil samples for *phoD* and *phoX*, respectively (Chapters 2, 3, 4 and 5). *phoD* and *phoX* were widely spread in bacteria compared to archaea and/or fungi (Figure 6.1). *phoD*- and *phoX*-harboring communities were hypothesized to have a similar composition. However, the *phoD*- and *phoX*-harboring communities had only some common phyla, namely *Acidobacteria, Actinobacteria, Planctomycetes* and *Proteobacteria*, with *Proteobacteria* being pre-dominant in most soils. In more detail, *phoD* was abundant in *Alpha*- and

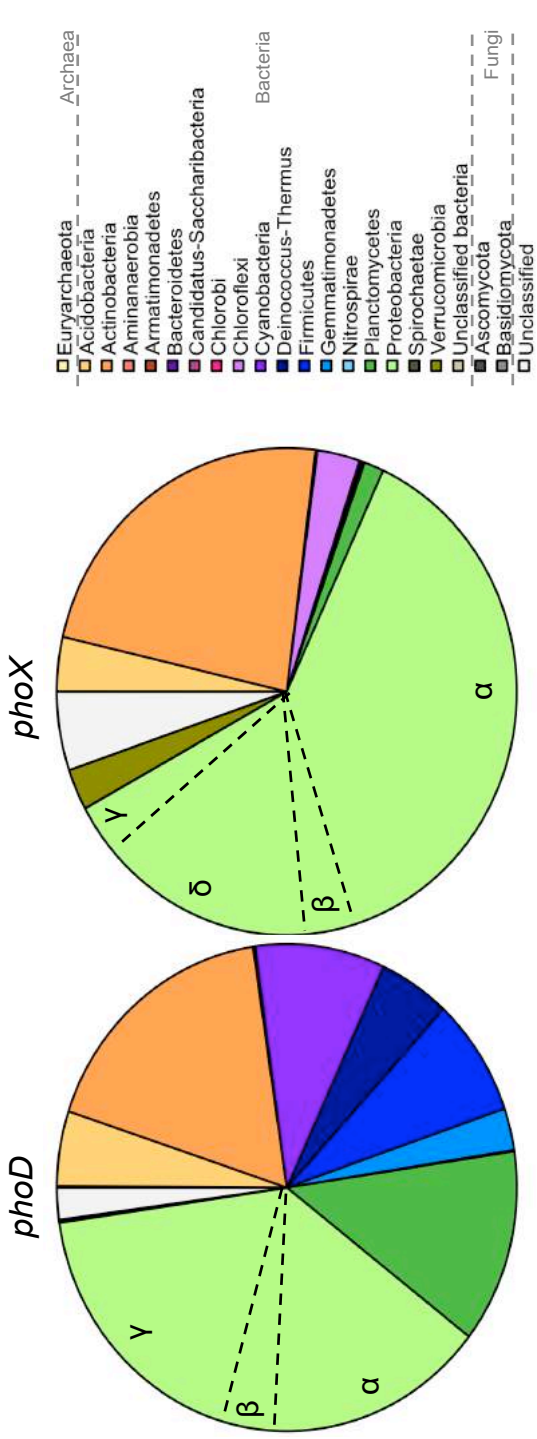


Figure 6.1. Mean relative abundance of each phylum across all studied soil samples of *phoD* and *phoX*. Dashed lines and Greek letters indicate the proportion of *Alpha-*(α), *Beta-* (β) and *Gamma-* (γ) *Proteobacteria* in the *phoD*- and *phoX*-harboring community and of *Delta-*(δ)*Proteobacteria* in the *phoX*-harboring community.

Gamma-Proteobacteria, while *phoX* was mainly found in *Alpha-Proteobacteria* (Figure 6.1). Other abundant *phoD*-harboring phyla included *Cyanobacteria, Deinococcus-Thermus, Firmicutes* and *Gemmatimonadetes* while other *phoX*-harboring phyla were *Chloroflexi* and *Verrucomicrobia*.

In soil, *phoD* and *phoX* were found in the common bacterial phyla that compose the total bacterial community such as *Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Gemmatimonadetes* and *Proteobacteria* (Lauber et al. 2009). This suggests that the *phoD*- and *phoX*-harboring communities are tightly linked to the total bacterial community composition. Additionally, the *phoD* gene was found in the two most common fungal phyla in soil, *Ascomycetes* and *Basidiomycetes*, which represent 70 to 98% of the fungal community (O'Brien et al. 2005). Moreover, this is the first time that the *phoD* and *phoX* genes are shown to occur in archaea.

The *phoD*-harboring community has been reported to be mainly composed of *Acidobacteria, Actinobacteria, Cyanobacteria* and *Proteobacteria* in a long-term fertilization trial on a humic Gleysol with a sandy loam texture (Tan et al. 2013) and of *Planctomycetes* and *Proteobacteria* in the ocean (Luo et al. 2009). In a metagenomic study that included a P-rich and a P-limited forest soil, dominant *phoD*-harboring bacterial phyla have been reported to consist of mainly *Proteobacteria* but also of *Acidobacteria* and *Actinobacteria* (Bergkemper et al. 2015). In more details, *Rhizobiales* were dominant in the P-rich soil, while *Actinomycetales*, *Acidobacteriales* and *Solibacterales* were dominant the P-limited forest soil. The main fungal phyla harboring acid and alkaline phosphatase genes was *Ascomycetes*.

Bergkemper et al. (2015) also showed that *phoD* was more abundant than *phoA*. The *phoX* gene was not amplified using shot-gun metagenomic on a P-limited and a P-rich forest soils (Bergkemper et al. 2015), showing that *phoX* is less abundant in forest soils than the *phoA* or *phoD* gene. The *phoX*-harboring community has been reported to be mainly composed of *Cyanobacteria* and *Proteobacteria* in fresh water ecosystems (Dai et al. 2014) and in the ocean (Luo et al. 2009). However, in this study, the *phoX* gene was

amplified in many forest soils, suggesting that the shot-gun metagenomic approach may not be sensitive enough to study the *phoX*-harboring community in certain soils.

Our results are in line with studies of Bergkemper et al. (2015), (Dai et al. 2014) and Luo et al. (2009), showing that a large fraction of the *phoD*- and *phoX*-harboring communities consists of *Proteobacteria*, which are known to be one of the most diverse bacterial phyla to harbor a substantial physiological diversity, such as heterotrophs, lithotrophs and phototrophs, and to be abundant in both aquatic and terrestrial ecosystems (Philippot et al. 2010). Besides *Proteobacteria*, for the *phoD*- as well as for the *phoX*-harboring community, the dominant phyla were different between aquatic and terrestrial ecosystems. These differences suggest that aquatic and terrestrial ecosystems consist of different dominant *phoD*- and *phoX*-harboring microorganisms and that terrestrial ecosystems harbor a higher diversity of *phoD*- and *phoX*-harboring microorganisms than aquatic ecosystems. Our results suggest that *phoD* and *phoX* are widely spread in the bacterial kingdom and are both found in high frequency in *Actinobacteria, Planctomycetes* and *Proteobacteria*. This supports our first hypothesized saying that *phoD* and *phoX* are on the whole found in similar phyla.

6.3.2 Active versus total *phoD*-harboring microorganisms

In Chapter 5, both the total and active *phoD*-harboring community compositions were investigated as affected by phosphate depletion and pH in a long-term P fertilization trial in grassland characterized by a pH gradient on the site. The key *phoD*-harboring microorganisms in the total and active communities remained the same between P fertilized and non-P fertilized treatments and across different pH levels and included *Acidobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria, Deinococcus-Thermus, Gemmatimonadetes, Planctomycetes* and *Proteobacteria* (Figure 6.2). These results are in disagreement with our general hypothesis stating that only some of the *phoD*-harboring phyla composing the total community are active.

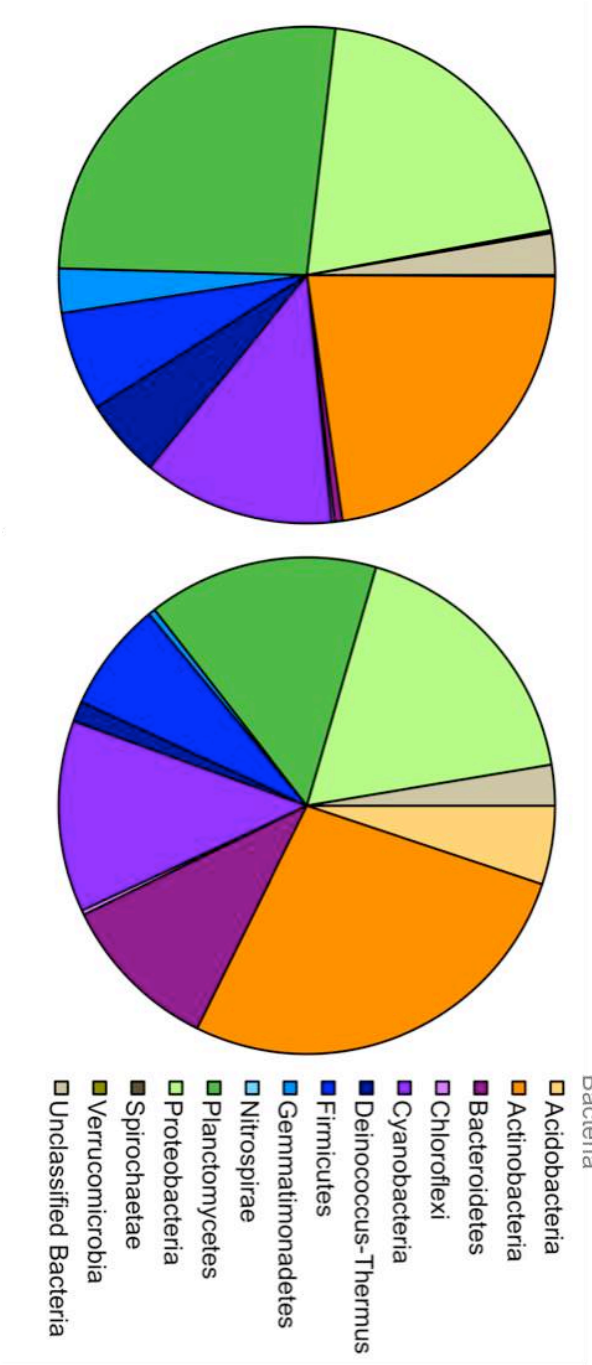


Figure 6.2. Mean relative abundances of the total and active *phoD*-harboring phyla observed in soil samples of Chapter 5.

In more details, the relative abundance of *Actinobacteria, Cyanobacteria* and *Proteobacteria* remained similar between the total and active *phoD*-harboring communities. However, the mean relative abundance of some other phyla differed greatly between the total and active *phoD*-harboring communities. For example, *Acidobacteria* and *Bacteroidetes* accounted for only 0.2 and 0.5% of the total *phoD*-harboring community, respectively, while they accounted for 5.0 and 10.3% of the active *phoD*-harboring community, respectively. On the other hand, *Firmicutes* represented 5.1% of the total and only 1.3% of the active *phoD*-harboring community.

Similarly to our results, Fraser et al. (2015b) observed the same dominant *phoD*-harboring phyla in both the total and active communities in soils fertilized with water-soluble phosphate or manure in a soil incubation study. Overall, however, the total and active *phoD*-harboring community compositions were different from the ones in our study. Dominant *phoD*-harboring phyla in their study were principally affiliated to *Actinobacteria, Cyanobacteria* and *Proteobacteria*, while a larger diversity of phyla was observed in ours. This discrepancy may be attributed to the different primer used. The primers applied by Fraser et al. (2015b) were designed by Sakurai et al. (2008) and have been reported to have an amplification bias towards *Alpha-Proteobacteria* (Tan et al. 2013) and to cover less diversity of *phoD* than our primers (See Chapter 2). Hence, the primers could explain the lower diversity of *phoD*-harboring microorganisms found in their study compared to ours.

The differences in relative abundance of certain phyla between the total and active *phoD*-harboring community observed in our study suggests that microorganisms respond differently to P limitation in soil. Hence, even microorganisms present only in low abundance in soil can express heavily phosphatase genes and thus represent key phosphatase producers. For example, in our study, the *phoD*-harboring *Acidobacteria* and *Bacteroidetes* were 25 and 21 times more abundant in the active than in the total *phoD*-harboring community, which is similar to the 21-fold up-regulation of the *phoD* gene reported in cultures of *Synechococcus* sp. OS-A and OS-B in a P deprivation study

(Adams et al. 2008). These values are a lot higher than the results reported in another culture-based P deprivation experiment which measured a 9.6 fold up-regulation of the *phoD* gene in *Aphanothece halophytica* (Kageyama et al. 2011). Together these studies and our results suggest that microorganisms induce the expression of *phoD* with different intensities, which may be due to differences in P concentration thresholds before inducing P starvation genes and to differences in the Pho regulon between microorganisms. The Pho regulon can consist of a different two-component signaling pathway from one bacteria to another, e.g. PhoR-PhoP in *Bacillus subtilis*, PnpR-PnpS in *Streptococcus pneumonia* and PhosS-PhosR in *Campylobacter jejuni* (Santos-Beneit 2015). Moreover, a two-component signaling pathway can either repress or induce phosphatase genes depending on the species (Santos-Beneit 2015). Additionally, the number of *phoD* and *phoX* gene copy per genome can vary from one species to another, reaching up to 9 and 5 copies, respectively (Chapter 2).

Additionally, a positive correlation between *phoD* gene abundance and potential alkaline phosphatase activity in soil suggests that PhoD contributes significantly to the total alkaline phosphatase activity (Fraser et al. 2015b). However, no correlation between the *phoD* transcript abundance and potential alkaline phosphatase activity was observed (Fraser et al. 2015b; Fraser et al. 2015a). Correlations between an enzyme-encoding gene abundance and the corresponding potential activity in soil have been often been observed, in contrast to correlations between an enzyme-encoding transcript abundance and its corresponding potential activity (Rocca et al. 2015). This can be explained by the fact that genes and enzymes are quite stable in soil and persist over time (Pettit et al. 1977), while transcripts have a short response time and a short half-life (Rocca et al. 2015). In our study, the abundances of *phoD* genes and transcripts were not quantified. However, our results showed that the *phoD*-harboring community composition was not correlated to the potential alkaline phosphatase activity (See Table 4.4 and Figure 5.4).

Our results suggest that soil microorganisms regulate *phoD* differently, and thus, it is essential to study the active *phoD*-harboring microorganisms to identify the potential key players involved in organic P hydrolysis in soil.

6.3.3 Evolutionary relationship between *phoD* and *phoX*

Among the three known alkaline phosphatase genes, namely *phoA*, *phoD* and *phoX*, *phoD* and *phoX* have often been reported in the same genome, while *phoA* seems to be rarely found in the same genomes as *phoD* or *phoX* (Luo et al. 2009; Zaheer et al. 2009). PhoX is believed to have similar function, properties and substrate affinity as PhoA (Luo et al. 2009; Zaheer et al. 2009), and thus would not be found in the same microorganisms due to their redundant function. The presence of *phoD* and *phoX* in many bacterial phyla, as well as *phoD* in archaea and fungi and *phoX* in archaea, can potentially reflect horizontal transfer of *phoD* and *phoX* between microorganisms in addition to vertical inheritance. It is possible that phosphatase genes have been horizontally transferred, i.e. via plasmids, between archaea, bacteria and fungi which has often been observed for genes involved in nutrient uptake (Brown and Doolittle 1997; Nelson et al. 1999; Pál et al. 2005; Ashby 2006). Supporting this assumption, the *phoX* gene has previously been found in a Pho operon on a plasmid in *Roseobacter denitrificans* (Sebastián and Ammerman 2009), which constitutes a common way to share genes among microorganisms (Brown and Doolittle 1997; Pál et al. 2005).

6.4 Relationships between *phoD/phoX* alkaline phosphatase genes and environmental factors

The relationships between the *phoD*/*phoX* phosphatase genes and the tested environmental variables were explored in 30 soils across 3 land-uses and 3 climate zones in Australia

and Switzerland (Chapter 4). *phoD* was in addition studied in the long-term fertilization trial in Watt (Switzerland) (Chapter 5). Our results are in agreement with our initial hypothesis stating that total *phoD*- and *phoX*-harboring community composition changes depending on environmental conditions. The tested environmental factors that governed the structure and composition of the *phoD*- and *phoX*-harboring communities were similar and included climate, land-use, soil group, total organic C, total N and total P, resin-extractable P, organic P and pH. However, the *phoD*-harboring community was only correlated with pH in the fertilization trial in Watt (Chapter 5) but not in the 30 soils study (Chapter 4). Although *phoD*- and *phoX*-harboring community structures and compositions were correlated to the same environmental factors, the level of significance of some of these correlations was different between the *phoD*- and *phoX*-harboring communities (Figure 6.3). The *phoD*-harboring community was strongly correlated with land-use and total organic C, while the *phoX*-harboring community was only weakly correlated to these factors. On the other end, while the *phoX*-harboring community was strongly correlated with available P, the *phoD*-harboring community was weakly affected by it.

Moreover, the relative abundance of some of the *phoD*- and the *phoX*-harboring phyla were correlated to certain environmental factors. For example, The relative abundance of *phoD*-harboring *Actinobacteria, Cyanobacteria* and *Firmicutes* increased to an optimum available P (P_{res}) between 25 and 50 mg P kg^{-1} soil, whereas the relative abundance of *phoD*-harboring *Proteobacteria* was lowest in this available P class (Figure 4.5a). The relative abundance of *phoD*-harboring *Planctomycetes* increased with available P and organic P (Figure 4.5a and 4.5b). In the *phoX*-harboring community, *phoX*-harboring *Chloroflexi* and *Verrumicrobia* was reached at a TOC of 10 to 20 g kg^{-1} soil (Figure 4.6a). Furthermore, the relative abundance of *phoX*-harboring *Actinobacteria* was greatest when TOC was below 10 g kg^{-1} soil (Figure 4.6a). Organic P was negatively correlated with the relative abundance of *phoX*-harboring *Actinobacteria* (Figure 4.6b) and *phoX*-harboring *Gemmatimonadetes*, which reached the highest relative abundance at a organic P of 250

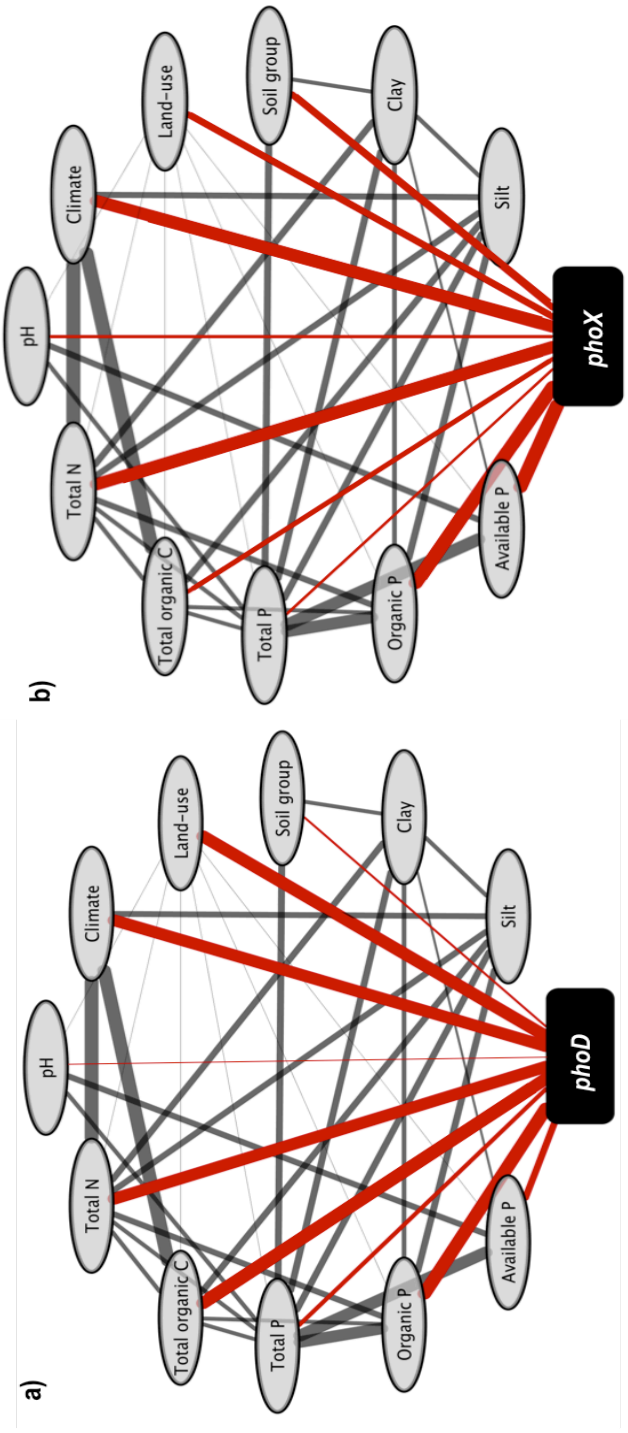


Figure 6.3. Conceptual summary of the correlations between the environmental factors (in gray), and between the environmental factors and the *phoD*-and the *phoX*-harboring community structure and composition (in red) analyzed in Chapter 4 and 5. Thickness of the line represents the level of significance from <0.05 (thinner lines) to <0.001 (thicker lines).

to 350 mg kg^{-1} soil (Figure 4.6b). Moreover, the land-use affected the relative abundance of some phyla in both the *phoD*- and *phoD*-harboring community. *phoD*-harboring *Cyanobacteria*, which were more abundant in grassland than in arable and forest soils, and of *phoD*-harboring *Firmicutes*, which were significantly lower in forest than in arable and grassland soils (Figure 4.5d). In the *phoX*-harboring community, the the relative abundance of *phoX*-harboring *Acidobacteria* was significantly higher in forest soils which were characterized by a lower pH than arable and grassland soils (Table 4.3 and Figure 4.6d). Nonetheless, a few *phoX*-harboring phyla clearly reacted to either land-use or pH.

The understanding and prediction of how the microbial community may respond to one or multiple environmental factors is one of the greatest challenges of microbial ecology today. To understand better the relationships between the microbial communities and the environment, it is essential to first assess the correlations between the environmental factors themselves. For example, in our studied soils, nutrient concentrations, e.g. total C, N and P, were largely interdependent and tightly linked to soil texture, land-use and climate (indicated in gray in Figure 6.3). Moreover, soil pH was strongly correlated to available P. Some of these correlations are well known, such as the link between pH and available P (Frossard et al. 1995; Hinsinger 2001) and the impact of land-use, e.g. via fertilization and grazing, on the concentration of soil C and N (Parton et al. 1987). Similarly, climate zones have a strong influence on some of the soil properties such as total C content (Cerling 1984; Burke et al. 1989), total N (Pastor and Post 1986) and biological activity (Monson et al. 2006). However, other correlations may be specific to the studied soil samples and have to be assessed case by case. The correlations between environmental factors create a complex network that makes it difficult to identify the environmental factors that ultimately shape the microbial community.

Nonetheless, our findings are supported by other studies reporting that available and total P are correlated with the *phoD*- and *phoX*-harboring community structure (Tan et al. 2013; Dai et al. 2014; Fraser et al. 2015b). Additionally, a study on Chilean Andisol pasture showed that P amendment did not affect the *phoD*-harboring community structure

unless applied together with N fertilizers, showing that both total P and N concentrations in soil are important determinants for the *phoD*-harboring community (Jorquera et al. 2014). As the synthesis of proteins requires energy and resources, particularly N, the availability of N in soil may limit phosphatase production (Sinsabaugh and Moorhead 1994; Allison et al. 2011). As a consequence, the N demand for the production of phosphatase may explain the strong correlation found between TN and the *phoD*- and *phoX*-harboring community structure and composition in our studies (Figure 6.3). In a metagenomic study, Bergkemper et al. (2015) compared P-limited and P-rich soils and showed that the P status can change the composition of the community harboring phosphatase genes. Our results are also in agreement with a study reporting a significant effect of crop management (including systems of monocropping, mixed culture and pre-cropping with legumes) on the *phoD*-harboring community structure (Wang et al. 2012a). To our knowledge, however, no study of *phoD* and *phoX* has yet been done on a global scale. Global-scale studies on total bacterial and fungal communities showed that climate, land-use and soil group are major environmental factors affecting the microbial community structure (Girvan et al. 2003; Drenovsky et al. 2004), followed by total N, total C, total P and pH for the bacterial community (Fierer and Jackson 2006; Lauber et al. 2008; Birkhofer et al. 2012) and total P for the fungal community (Lauber et al. 2008). *phoD* and *phoX* genes appear to be affected by similar environmental drivers as the total bacterial and fungal community. All in all, our results suggest that multiple environmental factors play a role in shaping the biogeographical pattern of *phoD*- and *phoX*-harboring community structure and composition and that the level of impact of these environmental factors differs for the *phoD*- and *phoX*-harboring communities.

6.5 Environmental perspective

Organic P is an important source of P for life in the environment (Paytan and McLaughlin 2007; Vitousek et al. 2010). It accounts for 20 to 65% of the total P in terrestrial

ecosystems (Harrison 1987) and 25 to 40% of the total P in aquatic ecosystems (Paytan and McLaughlin 2007). Phosphomonoesters, followed by phosphodiesters, represent often the dominant organic P fraction in both terrestrial and aquatic ecosystems (Harrison 1987; Kolowith et al. 2001). Organic P can be mineralized via non-biological processes, in which phosphoryl transfer occurs by loose transition states at appropriate pH, or by biological processes, in which phosphatases catalyse the organic P hydrolysis (Hengge 2005). As phosphoesters are stable molecules, phosphatases, mainly acid and alkaline phosphatases (See Section 1.2.2), are essentially responsible for the organic P mineralization in the environment. Phosphatase-labile organic P account for 35 to 89% of total extracted organic P in the topsoil (Jarosch et al. 2015) and for 19 and 50% of the organic P fraction in the euphotic zone in the ocean (Kobori and Taga 1979). Gross organic P mineralization rate ranges between 0.10 and 2.50 mg P kg^{-1} day^{-1} in soil (Oehl et al. 2004; Bünemann et al. 2007; Bünemann 2015) and between 0.15 and 0.33 mg P L^{-1} day^{-1} in lakes and estuaries (Chao et al. 2006; Shen 2006), showing that organic P mineralization is a valuable process to replenish the available P pool. Moreover, phosphatase-producing bacteria are found everywhere in the environment. They account for 46 to 83% of the bacteria in soil (Zimmerman et al. 2013) and for 40 to 50% of the heterotrophic bacteria in the ocean (Kobori and Taga 1979). Evidence shows that phosphatases, particularly alkaline phosphatases, found in terrestrial and aquatic ecosystems are principally of microbial origins (Jansson et al. 1988; Plante 2007). Moreover, some phosphatases are rather stable and persist in the environment. They have been reported to persist 6 days in lakes (Jansson et al. 1988) and 2 weeks in the ocean (Baltar et al. 2013) and at least a month (Pettit et al. 1977; George et al. 2005), up to several months to a year as some studies suggest (Li et al. 1998; Quiquampoix and Mousain 2005), in the soil.

Organisms such as microorganisms and plants have evolved to produce heat- and cold-stable phosphatases to access organic P in e.g. hot springs (Galperin and Jedrzejas 2001) and in arctic alpine altitude (Löffler et al. 2008), suggesting that organic P may be an important source of P in certain environments. For example, experimental addition

of phosphatases on desert soils showed that up to 87% of organic P is hydrolyzable if soils are wetted (Nadeau et al. 2007; Turner et al. 2003). Moreover, phosphatases can mineralize phosphate in excess of organisms' demand and, thus, contribute to P available for nutrition of other organisms such as plants (Weintraub and Schimel 2005; Richardson and Simpson 2011), and facilitate the coexistence of species (McKane et al. 2002).

Additionally, pH varies substantially in terrestrial ecosystems, while external aquatic ecosystems such as lakes and oceans usually have an alkaline pH optimum (Takahashi et al. 2014). This suggests that alkaline phosphatases such as PhoD and PhoX may play a larger role in aquatic than in terrestrial ecosystems. In contrast to most acid phosphatases which are mostly constitutively produced in bacteria, most alkaline phosphatases are up-regulated in P-limited conditions (Toh-e et al. 1973; Apel et al. 2007; Beazley et al. 2011). It has been suggested that the level of alkaline phosphatase activity may be used as an indicator of phosphorus deficiency in aquatic ecosystems (Vidal et al. 2003). However, the facts that the regulation of the gene expression of phosphatases can vary greatly between organism (Beardall et al. 2001) and that phosphatase production may also indicate a limitation in biomass but not necessarily a limitation in P (Dyhrman and Ruttenberg 2006) complicate the interpretation of alkaline phosphatase activity as indicator of P stress and suggest the use of multiple indicators to evaluate the state of P deficiency of a ecosystem. Furthermore, alkaline phosphatases such as PhoX have been found in P-rich aquatic environments such as Lake Taihu (Dai et al. 2014) and the Chesapeake Bay (Sebastián and Ammerman 2009). Diversity of *phoX* has been reported to be higher in hypereutrophic than mesotrophic regions of Lake Taihu, while the abundance of *phoX* was higher in the mesotrophic than hypereutrophic regions (Dai et al. 2014). These studies suggest that alkaline phosphatase is produced not only under P-limiting but also under P-rich conditions. Microorganisms that can produce alkaline phosphatase and, hence, access more P ressources, may be favoured also in P-rich conditions.

6.6 Agronomic perspective

Agricultural production is defined as the result of the transfer of nutrients from the soil to plants. At the beginning of the XIXth century, all possible materials known to increase soil fertility (e.g.: ashes, algae, wool wastes, household wastes, guano) were used to increase agricultural production (Frossard et al. 2009). By the middle of the XIXth century, the need for fertilizers was recognized and actively sought. England first started to use artificial fertilizers such as nitrate of soda and superphosphates to increase agricultural productivity (Frossard et al. 2009). In the 1960s, considerable progresses was made in soil science, plant nutrition and plant breeding, which resulted in the elaboration of new fertilization strategies. This marked the start of the "Green revolution", which dramatically improved agricultural production principally in the third world countries (Wharton 1969).

In the present context of increasing global food demand, improving nutrient use efficiency in plants while decreasing fertilizer inputs is a major challenge in agriculture. Maintenance of a sufficient level of available P in soil is essential to sustain productivity and is commonly reached by applying water-soluble phosphate fertilizer. Switzerland alone imports annually 16,508 tonnes of phosphorus, from which 70% is used in agriculture as feeds or fertilizers[1]. However, between 25 and 40% of the P applied is taken up by the crop, while the rest is fixed or bound onto mineral and organic surfaces (Schefe et al. 2015), leading to high organic and inorganic P stocks in soil (Syers et al. 2008) which can become only slowly available for the plants in the long-term (Frossard et al. 2014). Moreover, the overuse of phosphate fertilizer such as water-soluble phosphate or slurry can lead to P transfer into ground water, rivers and streams causing eutrophication of aquatic ecosystems and other environmental problems. Better exploration and exploitation of soil resources must be achieved in order to sustain our agroecosystems.

1 *Phosphorflüsse der Schweiz*, Schweizerbundesamt für die Umwelt, 2009, Bern, Switzerland

In this context, (i) plants able to take up P efficiently, (ii) prediction of plant-available P concentration in the soil over time and (iii) exploitation of the various P compounds in soil via soil biological activity may be essential to achieve sustainability in our agroecosystems (Oberson et al. 2006; Plassard et al. 2015). Within the same plant species, cultivars can exhibit genetic and phenotypic variation that can result in differences in their efficiency to acquire phosphorus from soil. The use of specific cultivars that have been selected via conventional plant breeding can improve P uptake efficiency in cropping systems (Ramaekers et al. 2010). Genetically modified plants, e.g: expressing phosphatases or long root hairs may also be a strategy to improve nutrient uptake efficiency. For example, the genetically modified *Arabidopsis thaliana* containing a phytase gene from *Aspergillus niger* may access more phosphate hydrolyzed from phytate in soil (Richardson et al. 2001). Genetically engineered *Solanum tuberosum* (potato) able to secrete stable phytase from trichoblasts was shown to accumulate 40% more P in leaves than wild-type plants when grown on quartz-loess-peat soil substrate (Zimmermann et al. 2003). Similarly, genetically engeneered *Trifolium subterraneum* (Subterranean clover) and *Nicotiana tabacum* (Tabacco) able to produce phytase were shown to be access P from phytate in sterile laboratory media. However, these genetically engineered lines were significantly less effective to utilize phytate as P sources in P-deficient soil collected from the field (George et al. 2004; George et al. 2005). It has been reported that phytase can be rapidly immobilized by adsorption in soil, inhibiting its activity for days and/or inactivating the enzyme (George et al. 2005). Therefore, further research on phosphatases and phosphatase-encoding genes is required. Additionally, optimizing the P supply based on available P and crop demand in space and time may reduce the use of fertilizer (Plassard et al. 2015). Finally, stimulating biological activity in the rhizosphere may improve the cycling of P in soil, and thus, plant nutrition (Frossard et al. 1995). Soil microorganisms play an important role in cycling P in soil by solubilizing inorganic P, mineralizing organic P and mobilizing P, replenishing the available P pool (Bünemann 2015). Moreover, it has been shown that P is cycled through microorganisms before being released into

the soil solution (Tamburini et al. 2012). Soil microorganisms and plants are in direct competition for readily available orthophosphate. However, the turnover time of the microbial biomass and the release rate is fast and thus microorganisms play an important role in the replenishment of the soil available P pool in the long term (Richardson et al. 1994; Oberson and Joner 2005). Moreover, based on laboratory experiments, it has been suggested that phosphatases released by microorganisms are more efficient at hydrolyzing organic P than phosphatases released by plants (Tarafdar et al. 2001). This difference is possibly be due to the co-production of organic acids such as malate, citrate, oxalate and phosphatases by microorganisms which may help the stabilization of the released orthophosphate ions in the soil solution. Nurturing the microbial activity in soil may, thus, improve P use efficiency and reduce the use of fertilizer in cropping systems. For example, crop management such as addition of manure and no till has been reported to increase the gross organic P mineralization rate (Oehl et al. 2004; Nannipieri et al. 2011). Additionally, phosphatase-producing organisms have been suggested to be used to coat seeds or as biofertilizers (Plassard and Dell 2010).

Each of these strategies represents a potential increase of nutrient use efficiency, and combined, a reduction of fertilizer applications. Together, these different strategies will optimize the use of the soil resources. Nonetheless, a deeper knowledge and further research on phosphatases and phosphatase-producing organisms in soil is required to fully understand to potential of phosphatases and to identify the important players that hydrolyze organic P in soil. A first step to acquire this knowledge includes the application of molecular tools such as the primers developed in this thesis.

7

MAIN CONCLUSIONS

In this thesis, new primer sets have been designed to target *phoD* and *phoX* in soil microorganisms and have been shown to amplify *phoD* and *phoX* with good coverage and specificity. *phoD* and *phoX* were found to be widely spread in the environment, from terrestrial to aquatic ecosystems, free-living and associated to various hosts. Using the newly-designed primers it was shown that

- the *phoD* gene was present in 1 archaeal, 13 bacterial and 2 fungal phyla, and the *phoX* gene in 1 archaeal and 16 bacterial phyla,
- the *phoD* and *phoX* gene were found in part in common phyla such as *Actinobacteria, Proteobacteria* and *Planctomycetes*,
- the total and active *phoD*-harboring communities were composed of similar phyla that at times differed greatly in relative abundances,
- despite differences in environmental factors, dominant phyla composing the *phoD*-harboring community were generally similar in all soil samples, while the *phoX*-harboring community composition differed substantially between the soil samples,
- multiple environmental factors were strongly correlated to the *phoD*- and *phoX*-harboring community structure and composition and these included climate, soil group, land-use, pH and soil nutrient concentrations,
- in more detail, our results showed that phosphate depletion affected the composition and structure of the active *phoD*-harboring community only, while soil pH impacted that of both the total and active *phoD*-harboring community.

8

OUTLOOK

The complexity of belowground communities, their interactions with plants and environmental factors and the link between total community and functioning is difficult to explore. In this study, molecular tools represented a valuable and innovative way to identify and study the unseen and immeasurable *phoD*- and *phoX*-harboring communities. Further studies are required to improve our knowledge on the key phosphatase-producing microorganisms in different environments. Moreover, many additional phosphatases (Table 1.1) are produced by microorganisms and plants to access P from organic P in soil. To date, another alkaline phosphatase gene, the *phoA* gene, and a few acid phosphatase genes such as the *acpA* acid phosphatase have been described. The use of specific primers represents a great and innovative tool to improve our knowledge on phosphatase genes. To allow the design of further primers, there is a need to create a reliable classification with naming standards to provide consistent annotation from the gene to the enzyme. Currently, classifications of enzymes and enzyme-encoding genes are nearly completely unrelated and bridges between the two classifications can be made only in a few cases, making the annotation in databases unclear and the design of primers even more difficult. Additionally, there is a lack of tools to efficiently analyze large sets of data of non-model organisms and not well-described genes.

Another challenge to fully understand phosphatases in soil is to link the different levels of the phosphatase synthesis from the gene to its catalysis (Figure 8.1). For example, sequencing and qPCR would allow to identify and quantify the key organisms that harbor and express these genes. Once transcripts have been translated and the amino acids properly folded, phosphatase enzymes can be extracted and analyzed using proteomics techniques. After the enzyme is activated with the right co-factor, potential phosphatase activity can be measured, e.g. using fluorescent assays. Moreover, information on the actual phosphatase processes can be obtained by using δ^{18}O-P as isotopic tracer (Tamburini et al. 2014) and organic P hydrolysis rate can determined e.g. using ^{33}P dilution techniques.

Finally, mineralization rate and P pools can be modelled (Bünemann 2015; Frossard et al. 1996) (Figure 8.1). The combination of multiple techniques would enable us to better understand the importance of phosphatases and phosphatase-producing organisms in soil, and the conditions required for phosphatase production (e.g. soil moisture, total C and N, etc.), starting from the gene.

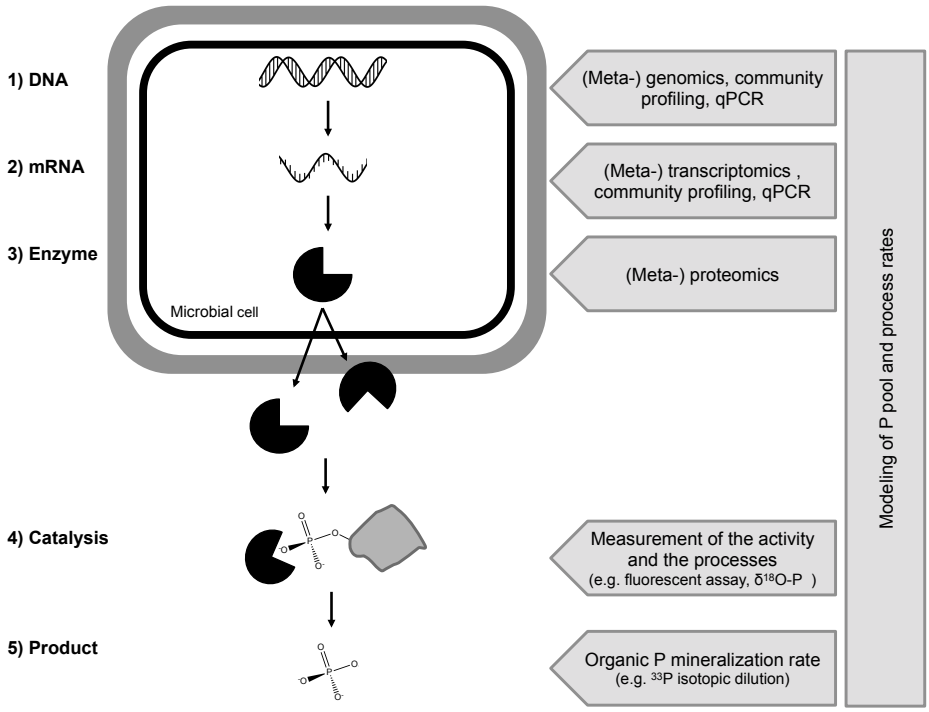


Figure 8.1. The steps of phosphatase synthesis from DNA to enzyme with techniques to study phosphatases at each step.

9

APPENDICES

A.1 *phoD* alkaline phosphatase gene diversity in soil

Published as:

Ragot SA, Kertesz MA, Bünemann EK (2015) *phoD* alkaline phosphatase gene diversity in soil. Applied and Environmental Microbiology 81, 7281-7289.

AEM
Journals.ASM.org

phoD Alkaline Phosphatase Gene Diversity in Soil

Sabine A. Ragot,[a] Michael A. Kertesz,[b] Else K. Bünemann[a]

Institute of Agricultural Sciences, ETH Zurich, Lindau, Switzerland[a]; Department of Environmental Sciences, Faculty of Agriculture and Environment, University of Sydney, Sydney, New South Wales, Australia[b]

Phosphatase enzymes are responsible for much of the recycling of organic phosphorus in soils. The PhoD alkaline phosphatase takes part in this process by hydrolyzing a range of organic phosphoesters. We analyzed the taxonomic and environmental distribution of *phoD* genes using whole-genome and metagenome databases. *phoD* alkaline phosphatase was found to be spread across 20 bacterial phyla and was ubiquitous in the environment, with the greatest abundance in soil. To study the great diversity of *phoD*, we developed a new set of primers which targets *phoD* genes in soil. The primer set was validated by 454 sequencing of six soils collected from two continents with different climates and soil properties and was compared to previously published primers. Up to 685 different *phoD* operational taxonomic units were found in each soil, which was 7 times higher than with previously published primers. The new primers amplified sequences belonging to 13 phyla, including 71 families. The most prevalent *phoD* genes identified in these soils were affiliated with the orders *Actinomycetales* (13 to 35%), *Bacillales* (1 to 29%), *Gloeobacterales* (1 to 18%), *Rhizobiales* (18 to 27%), and *Pseudomonadales* (0 to 22%). The primers also amplified *phoD* genes from additional orders, including *Burkholderiales*, *Caulobacterales*, *Deinococcales*, *Planctomycetales*, and *Xanthomonadales*, which represented the major differences in *phoD* composition between samples, highlighting the singularity of each community. Additionally, the *phoD* bacterial community structure was strongly related to soil pH, which varied between 4.2 and 6.8. These primers reveal the diversity of *phoD* in soil and represent a valuable tool for the study of *phoD* alkaline phosphatase in environmental samples.

Phosphorus (P) is an essential macronutrient for all living cells (1). Despite its relative abundance in soils, P is one of the main limiting nutrients for terrestrial organisms (2). P is present in organic and inorganic forms in soil, but only the inorganic orthophosphate ions in soil solutions are readily available for plants (3). To sustain crop productivity, large amounts of P fertilizers are therefore used in agriculture, both as inorganic fertilizers (e.g., triple super phosphate) and organic fertilizers (e.g., manure). After application, some of the inorganic P is rapidly taken up by plants and microorganisms, while the remaining P is immobilized as insoluble and bound P forms in the soil. Microorganisms can access and recycle P from these recalcitrant P forms by solubilization of inorganic P and by mineralization of organic P via enzymatic processes mediated primarily by phosphatases, which hydrolyze the orthophosphate group from organic compounds (3). When facing P scarcity, microorganisms upregulate expression of functional genes coding for phosphatases (phosphomonoesterases, phosphodiesterases, phytases), high-affinity phosphate transporters, and enzymes for phosphonate utilization, which together constitute the Pho regulon (4). The phosphomonoesters which are hydrolyzed by phosphatases are generally the dominant fraction of organic P and can represent up to 90% of the organic P in soil (3).

Prokaryotic alkaline phosphatases have been grouped into three distinct families, PhoA, PhoD, and PhoX (5–7), which are classified in COG1785, COG3540, and COG3211, respectively, of the Cluster of Orthologous Groups (COG) categorization. PhoA was the first alkaline phosphatase to be characterized. It is a homodimeric enzyme that hydrolyzes phosphomonoesters and is activated by Mg^{2+} and Zn^{2+} (7). PhoD and PhoX are monomeric enzymes that hydrolyze both phosphomonoesters and phosphodiesters and are activated by Ca^{2+} (5, 6). Enzymes of all three families are predominantly periplasmic, membrane bound, or extracellular (8). PhoD and PhoX are exported by the twin-arginine translocation pathway (5, 6), while PhoA is secreted via the Sec protein translocation pathway (9). There is high sequence variability in the PhoA, PhoD, and PhoX proteins, not only between the families but also within each family (5, 9). PhoD is widespread in both terrestrial and aquatic ecosystems (8, 10).

Until recently, our knowledge of the phosphatase-encoding genes in prokaryotes was based on traditional culture-dependent methods. Advances in culture-independent techniques have provided new tools for the study of microbial communities in the environment. The first functional gene probes to target alkaline phosphatase genes were the primers developed by Sakurai et al. named ALPS primers (11). They were based on phosphatase gene sequences from seven isolates and first used to examine the different soil alkaline phosphatase community structures resulting from mineral and organic fertilization. Alkaline phosphatase genes belonging to the *Actinobacteria*, *Alpha-*, *Beta-*, and *Gammaproteobacteria*, and *Cyanobacteria* classes were identified by clon-

Received 2 June 2015 **Accepted** 1 August 2015

Accepted manuscript posted online 7 August 2015

Citation Ragot SA, Kertesz MA, Bünemann EK. 2015. *phoD* alkaline phosphatase gene diversity in soil. Appl Environ Microbiol 81:7281–7289. doi:10.1128/AEM.01823-15.

Editor: G. Voordouw

Address correspondence to Sabine A. Ragot, sabine.ragot@usys.ethz.ch.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01823-15.

doi:10.1128/AEM.01823-15

ing, giving the first insight into alkaline phosphatase diversity in soil (11).

Subsequently, the ALPS primers were demonstrated to be specific to the *phoD* alkaline phosphatase gene (10). They were used to assess alkaline phosphatase gene diversity and structure in several soils by PCR-denaturing gradient gel electrophoresis (DGGE) (12–15) and by 454 sequencing (10, 16). These studies showed that crop management, application of organic and conventional fertilizers, and vegetation all affect the *phoD* alkaline phosphatase gene diversity. Tan et al. (10) examined the effects of three mineral P fertilization intensities (zero, medium, and high input) in grassland soil on the composition and diversity of alkaline phosphatase and found a change in the *phoD* bacterial community compositions between unfertilized and fertilized treatments, with the dominant *phoD* alkaline phosphatase genes affiliated with *Alpha-* and *Gammaproteobacteria*, *Actinobacteria*, and *Cyanobacteria*. However, they pointed out that the ALPS primers are likely to have an amplification bias, resulting in an overrepresentation of *Alphaproteobacteria*, and that new primers are therefore required to provide better coverage of the *phoD* diversity.

In this study, we assessed the diversity and environmental distribution of the *phoD* gene based on current genome and metagenome databases, and we present a new set of improved primers which targets the large diversity of *phoD* genes in soil microorganisms. These primers can be used as a tool both to identify PhoD-producing bacteria and to study *phoD* bacterial community diversity and composition in the environment. The newly designed primers were tested in a gene-targeted metagenomic approach using 454 sequencing in a range of soils collected from two continents with different climates and soil properties. Finally, we compared them to the previously published ALPS primers (11), using the same samples and methodology.

MATERIALS AND METHODS

Taxonomic and environmental distribution of *phoD* alkaline phosphatase genes across microbial genomes and metagenomes. The distribution of *phoD* genes was assessed using the Integrated Microbial Genomes and Metagenomes (IMG/M) database, a dedicated system for annotation of whole genomes and metagenomes (17). Draft and complete genome data sets were used to evaluate the distribution of *phoD* across kingdoms and phyla, and metagenome data sets were used to evaluate the prevalence of *phoD* in the environment (data accessed on 13 July 2015). Metagenome data sets were categorized as "air," "engineered and waste" (bioreactor and waste treatment), "extreme environments" (saline, alkaline, hot spring, brine, and black smokers), "fresh water," "marine environment," "plant-associated" (leaves and wood), "animal-associated" (associated with humans, arthropods, molluscs, and sponges), and "soil" (rhizosphere and bulk soil). These categories were chosen based on the environment-type classification of the IMG/M database. The relative abundance of *phoD* gene counts per environment type was calculated as the gene count number normalized by the total number of bases sequenced per metagenome data set.

Soil sampling and general soil characteristics. Four grassland soils were collected in Australia in July 2013 (samples 1 to 4 [S1 to S4]), and two grassland soils were sampled in Switzerland in September 2012 (S5 and S6) (Table 1). These represent a broad range of soil types, vegetation, and climatic conditions, varying from hot semiarid to continental temperate climates. At each site, five soil cores from the top 5 cm were randomly collected and homogenized by sieving (4 mm). A subsample was stored at −80°C for molecular analysis. The remaining composite soil was air dried and used to determine the basic soil properties, including pH, texture, and total carbon (C) and P. Methods used to determine the soil properties are

TABLE 1 Description of the grassland soils S1 to S6, with location, geographical coordinates, climate, soil type, vegetation, pH, texture, and total C and P

Sample	Site	Geographical coordinates	Climate (climate category)[a]	Soil type[b]	Vegetation	pH ($CaCl_2$)[c]	Texture (% clay, % silt, % sand)[d]	Total C (g kg^{-1} soil)[e]	Total P (mg kg^{-1} soil)[f]
S1	Kia-Ora (Australia)	34°48′18″S, 148°35′00″E	Warm temperate climate, fully humid with warm summer (Cfb)	Planosol	*Microlaena stipoides*, *Austrodanthonia* spp., *Elymus scaber*, *Bothriochloa macra*, *Austrostipa* spp.	4.2 ± 0.3	14, 28, 58	21.0 ± 0.8	221 ± 8
S2	Narrabri (Australia)	30°15′14″S, 149°51′53″E	Warm temperate climate, fully humid with warm summer or with hot summer (Cfa)	Planosol	*Chrysocephalum* sp., *Themeda* sp., *Festuca arundinacea*	6.1 ± 0.0	38, 27, 35	23.7 ± 0.1	705 ± 13
S3	Nyngan (Australia)	31°25′52″S, 147°04′09″E	Arid climate, hot steppe (BSh)	Cambisol	Mixed grasses and dicot plants; clumpy cover, not a sward	4.7 ± 0.1	30, 33, 37	15.0 ± 0.3	466 ± 10
S4	Mutawintji (Australia)	31°16′19″S, 142°17′44″E	Arid climate, hot steppe (BSh)	Leptosol	*Chenopodium* sp., *Acacia* sp., *Astrebla* sp.	6.8 ± 0.1	12, 11, 77	5.0 ± 0.2	193 ± 11
S5	Watt (Switzerland)	47°25′45″N, 008°29′31″E	Warm temperate climate, fully humid with warm summer (Cfb)	Cambisol	*Arrhenaterion elatioris*	5.0 ± 0.1	30, 33, 37	27.5 ± 0.1	613 ± 33
S6	Watt (Switzerland)	47°25′45″N, 008°29′31″E	Warm temperate climate, fully humid with warm summer (Cfb)	Cambisol	*Arrhenaterion elatioris*	6.1 ± 0.3	30, 33, 37	34.4 ± 0.4	703 ± 39

[a] Köppen-Geiger climate classification. Climate categories are described further in a paper by Kottek et al. (51).
[b] World Reference Base for Soil Resources (52).
[c] Measured in a soil suspension in 0.01 M $CaCl_2$ with a 1:2.5 mass/volume ratio using a Benchtop pH/ISE 720A (Orion Research Inc., Jacksonville, FL).
[d] Determined by a commercial soil analysis lab (Soil Conseil, Nyon, Switzerland).
[e] Measured on dry and ground soil using a CNS analyzer (Thermo-Finnigan).
[f] Determined by wet digestion with H_2O_2-H_2SO_4 (53) and measured with malachite green at 610 nm (54).

described in Table 1. The sampled soils covered a range of textures, with clay contents varying between 12 and 38%. Soil pH ranged between 4.2 and 6.8. Total C varied between 5 and 34 g kg^{-1} soil, and total P varied between 193 and 705 mg kg^{-1} soil. The vegetation densities were similar at sampling sites S5 and S6 but very different at the other sites, ranging from dense to scarce, depending on the location.

DNA extraction from soil. All DNA samples were extracted in duplicate. Nucleic acids were extracted from the Australian samples using a DNA PowerSoil isolation kit (Mo Bio, Carlsbad, CA, USA), according to the manufacturer's instructions, with an initial bead-beating step of 2 cycles of 3 min at 30 Hz using a TissueLyser II (Qiagen, CA). Nucleic acids were extracted from the Swiss samples (2 g of frozen soil) using an RNA PowerSoil isolation kit (Mo Bio) according to the manufacturer's instructions, with an additional homogenizing step using an Omni Bead Ruptor homogenizer (Omni International, Kennesaw, GA) (2.8-mm zirconium beads for 1 min at 5 m s^{-1}) prior to isolation. DNA was eluted from the RNA/DNA capture column using 4 ml of DNA elution solution (1 M NaCl, 50 mM morpholinepropanesulfonic acid [MOPS], 15% isopropanol [pH 7]). DNA was precipitated using isopropanol and resuspended in diethyl pyrocarbonate (DEPC)-treated H_2O. Only the DNA extracts were used in this study.

Primer design and *in silico* testing. Gene sequences annotated as *phoD* and/or associated with COG3540 (Clusters of Orthologous Groups; http://www.ncbi.nlm.nih.gov/COG/), which corresponds to *phoD* alkaline phosphatase, were retrieved from the European Nucleotide Archive (ENA) and UniProt Knowledgebase (UniProtKB) databases. They were then clustered at 97% similarity using CD-HIT (18), resulting in a total of 315 sequences used as references for the primer design (see the list in Table S1 and the taxonomic tree in Fig. S1 in the supplemental material). The reference sequences were affiliated with 11 phyla, including *Actinobacteria* (59 sequences), *Bacteroidetes* (22 sequences), *Cyanobacteria* (22 sequences), *Deinococcus-Thermus* (2 sequences), *Ignavibacteriae* (1 sequence), *Firmicutes* (13 sequences), *Gemmatimonadetes* (1 sequence), *Spirochaetes* (16 sequences), *Planctomycetes* (4 sequences), *Proteobacteria* (173 sequences), and *Verrucomicrobia* (2 sequences).

The gene sequences were aligned using MUSCLE (19), and the alignment was manually reviewed by comparison with the aligned translated sequences, using Geneious 6.1.2 (Biomatters, Australia) and the alignment of the COG3540 group available on the NCBI website (Conserved Domain Protein Family, http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) as the amino acid reference alignment. The most suitable conserved regions for primer design were identified using PrimerProspector (20). Forward and reverse candidate primers were then manually designed to reach the maximum coverage of the reference sequences. Candidate primers were paired to target an amplicon length of 250 to 500 bp, which represents the best compromise length for next-generation sequencing and quantitative PCR studies. They were then tested *in silico* using De-MetaST-BLAST (21) to identify potential primer pairs with an appropriate product size and coverage of the reference sequences.

Optimization and validation of *phoD*-targeting primers. Candidate primers (21 forward primers and 23 reverse primers) were tested in a gradient PCR using a mixture of soil genomic DNA (S5 and S6) (Table 1) as the template. PCRs were performed in a 25-μl volume containing 1× MyTaq reaction buffer (including $MgCl_2$ and deoxynucleoside triphosphates [dNTPs]), 0.5 μM each primer, and 0.6 U of MyTaq polymerase (Bioline, NSW, Australia) with 1 to 2 ng DNA as the template in an S1000 thermocycler (Bio-Rad Laboratories, CA). The amplification reaction included an initial denaturation step of 5 min at 95°C, followed by 35 cycles of a denaturation step of 30 s at 95°C and an annealing step of 30 s at the calculated annealing temperature of each candidate primer pair (gradient of ±3°C), and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. Amplicon size and intensity and the presence of primer dimers were assessed visually after electrophoresis on a 1.5% (wt/vol) agarose gel and staining with ethidium bromide.

The amplicon specificity was evaluated for selected primer pairs by cloning and sequencing. The PCR products were ligated at 4°C overnight using pGEM-T vector systems (Promega, Madison, WI) and transformed into chemically competent *E. coli* cells [α-select; F^- *deoR endA1 recA1 relA1 gyrA96 hsdR17* (r_k^- m_k^+) *supE44 thi-1 phoA* Δ(*lacZYA-argF*)*U169* Φ80*lacZ*ΔM15 λ^-] according to the manufacturer's instructions (Bioline). Restriction fragment length polymorphism (RFLP) profiling of clones with the expected insert size was done using HhaI (0.2 U/μl for 3 h at 37°C; Promega), and profiles were visualized by electrophoresis on a 2% (wt/vol) agarose gel. Representative inserts of unique RFLP profiles were then sequenced (Macrogen Inc., Seoul, South Korea). The resulting sequences were used to evaluate the coverage and specificity of the candidate primer pairs using BLAST (22).

Amplicon diversity was examined for three candidate primer pairs by 454 GS-FLX+ sequencing (Roche 454 Life Sciences, Branford, CT) using barcoded primers. The barcoded primer design, sequencing, and initial quality filtering were performed by Research and Testing Laboratory (Lubbock, TX) using standard protocols. Briefly, sequences with a quality score of <25 were trimmed, and chimeras were removed using USEARCH, with clustering at a 4% divergence (23). Denoising was performed with the Research and Testing Denoise algorithm, which uses the nonchimeric sequences and the quality scores to create consensus clusters from aligned sequences. Within each cluster, the probability of prevalence of each nucleotide was calculated, and a quality score was generated, which was then used to remove noise from the data set.

The primer pair *phoD*-F733 (5′-TGGGAYGATCAYGARGT-3′)/*phoD*-R1083 (5′-CTGSGCSAKSACRTTCCA-3′) provided the highest *phoD* diversity and coverage (numbers indicate the respective positions in the reference *phoD* gene of *Mesorhizobium loti* MAFF303099). *phoD*-F733 anneals to the conserved region that consists of the amino acid residues WDDHE, which contribute to the coordination of two Ca^{2+} cofactors (24). In addition, the fragment targeted by *phoD*-F733/*phoD*-R18083 includes two conserved arginine residues. Nevertheless, the variable part of the amplified region also allows a good identification of taxonomy. This primer pair was named PHOD and used further in this study.

454 sequencing using PHOD and ALPS primers. For comparative analysis of PHOD and ALPS primers ALPS-F730/ALPS-R110 (5′-CAGTGGGACGACCACGAGGT-3′/5′-GAGGCCGATCGGCATGTCG-3′) (11), *phoD* genes were amplified in pooled duplicate DNA extracts at a concentration of 20 ng μl^{-1} using the PCR conditions described above, with an annealing temperature at 58°C for PHOD primers and at 57°C for ALPS primers. Samples were then sequenced using 454 GS-FLX+ pyrosequencing (Roche) by Research and Testing Laboratory, with a resulting yield between 1,642 and 13,998 reads per library.

Sequence analysis. Sequencing data sets amplified by PHOD and ALPS primers were analyzed separately using mothur (25). Sequences were analyzed as nucleic acid sequences to keep the maximum information, allow accurate identification, and avoid artifacts due to frameshifts and errors during back-translation (26). After demultiplexing, reads containing ambiguities and mismatches with either the specific primers or the barcode were removed. Reads with an average quality score of <20 were then filtered out. The remaining reads were trimmed at 150 bp and 450 bp as the minimum and maximum lengths, respectively. Across all samples, 92% of the sequences had a length between 320 and 380 bp.

The resulting PHOD- and ALPS-amplified data sets were merged and aligned using the Needleman-Wunsch global alignment algorithm as implemented in mothur, using 6-mer searching and the aligned reference sequences as the template. The pairwise distance matrix was calculated from the alignment, and sequences were clustered using the k-furthest method as implemented in mothur, with a similarity cutoff at 75% to define the operational taxonomic units (OTUs), as calculated by Tan et al. (10). OTU matrices were normalized to the smallest library size using the normalized.shared command in mothur to allow comparison between samples. The relative abundance of each OTU was normalized by the total number of reads per sample. The normalized values were then rounded to the nearest integer. The taxonomy assignment was performed using

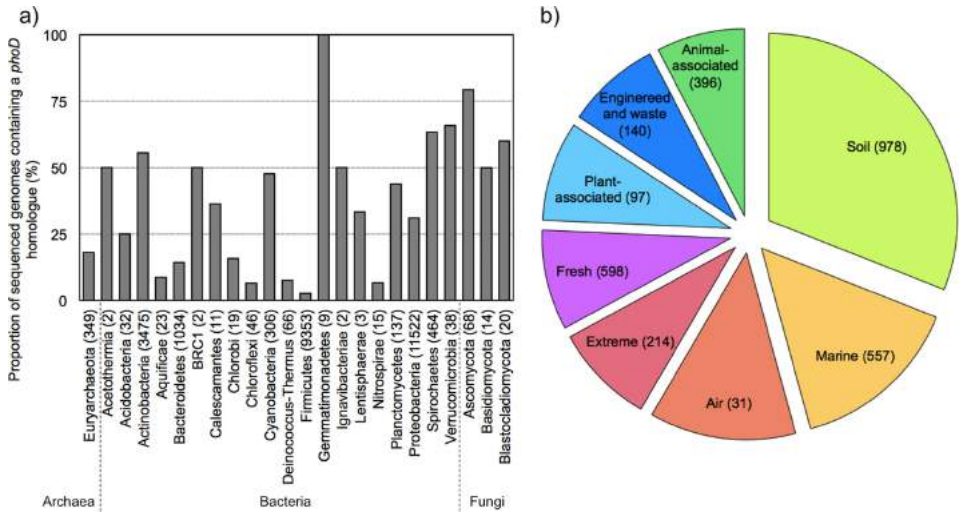


FIG 1 Current knowledge of the *phoD* gene in the IMG/M database. (a) Proportion of sequenced genomes containing a *phoD* homologue. The numbers in parentheses indicate the total number of sequenced genomes in each phylum. (b) Relative abundance of *phoD* genes in different types of environments (normalized as the number of *phoD* counts per number of bases sequenced per metagenome data set). The numbers in brackets indicate the number of metagenome data sets per environment type.

BLASTn in BLAST+ (27), with a minimum E value of 1e−8 to retrieve NCBI sequence identifiers (GI accession number). Subsequently, in-house Perl scripts were used to populate and query a MYSQL database containing the NCBI GI number and taxonomic lineage information (the scripts were written by Stefan Zoller, Genetic Diversity Centre, ETH Zurich, and are available on request).

Data analysis. Rarefaction curves were calculated and extrapolated to 5,000 reads to standardize the samples using EstimateS (version 9; http://purl.oclc.org/estimates). The unconditional variance was used to construct 95% confidence intervals for both interpolated and extrapolated values, which assumes that the reference sample represents a fraction of a larger but unmeasured community. Observed species richness (S_{obs}) based on the normalized library size, estimated species richness based on a library size of 5,000 reads (S_{est}), and the Chao1 species richness index (28) were calculated using EstimateS. Additionally, the Good's coverage (29) and the alpha diversity estimated by the Shannon-Wiener (H′) (30) index were calculated. Paired Student *t* tests were used to compare S_{obs}, S_{est}, Good's coverage, and H′ indices between samples.

Similarities between *phoD* bacterial community structures were tested using pairwise libshuff analysis as implemented in mothur with 1,000 iterations (31). Correlations between the community composition and environmental variables were tested by redundancy analysis (RDA), followed by an analysis of variance (ANOVA) on the RDA fit, and a variance partitioning analysis using the vegan package (vegan; Community Ecology Package, R package version 2.2-0; http://CRAN.R-project.org/package=vegan) in R version 2.15.0 (R Core Team, 2014; http://www.R-project.org). Prior to analysis, the measured environmental variables (clay and silt content, total C and P, and soil pH) were standardized using the Z-score method, and nominal variables (vegetation, climate, and soil type) were also included.

Nucleotide sequence accession number. The standard flowgram format (.sff) files were submitted to the European Nucleotide Archive (ENA) under the accession number ERP008947.

RESULTS AND DISCUSSION

Taxonomic distribution of *phoD* alkaline phosphatase gene. Our current knowledge of the taxonomic distribution of *phoD* was described based on the IMG/M database. A total of 63 archaeal, 6,469 bacterial, and 73 eukaryotic draft or complete genomes containing at least one copy of the *phoD* gene were found.

In bacteria, the *phoD* gene was spread across 20 phyla (Fig. 1a). More than half of the genomes of *Actinobacteria*, *Gemmatimonadetes*, *Spirochaetes*, and *Verrucomicrobia* contained at least one copy of the *phoD* gene. Among the *Proteobacteria*, the *phoD* gene occurred in 52, 30, and 34% of the *Alpha-*, *Beta-*, and *Gammaproteobacteria*, respectively. The number of *phoD* copies per genome varied between 1 and 9, but the majority of sequenced genomes (71%) carried only a single copy.

Although *phoD* is widespread across the bacterial phyla, it is important to note that the microbial genome sequence database contains the genomes of cultured strains almost exclusively, which creates a general bias in databases (32). *Proteobacteria* was the most recurrent phylum in the database, as the *Gammaproteobacteria* and more particularly the *Pseudomonas* genus are among the most intensively studied taxa (32) and thus are the genomes found most frequently in databases. Given the presence of the *phoD* gene in the less represented phyla, such as *Chloroflexi*, *Deinococcus-Thermus*, and *Planctomycetes*, *phoD*-targeting primers represent an important tool to study these less easily culturable phyla.

Additionally, *phoD* genes were found in archaea, affiliated almost entirely with *Euryarchaeota* (*Halobacteriaceae*), and in eukaryotes, mainly in *Ascomycetes*. Alkaline phosphatase activity in archaea has only rarely been reported, e.g., from extreme environments (33, 34), while in eukaryotes it has been reported in *Basidiomycetes* (35) and in eukaryotic phytoplanktonic cells (36); in mammals, it is widely used as an indicator for liver disease (37). However, alkaline phosphatase activity has not previously been associated with the *phoD* gene in these taxa.

Environmental distribution of *phoD* alkaline phosphatase—a meta-analysis. The prevalence of *phoD* in the environ-

TABLE 2 Data obtained with PHOD and ALPS primers based on normalized data[a]

Primer	Sample	No. of filtered reads	No. of unique reads	No. of reads after normalization	Species richness index			Good's coverage	H′	No. of:				
					S_{obs}	S_{est}	Chao1			Phyla	Classes	Orders	Families	Genera
PHOD	S1	1,915	1,763	1,088	290	685	684	0.83	4.6	10	15	20	30	37
	S2	2,170	1,820	963	201	293	303	0.91	3.9	10	14	18	29	39
	S3	3,090	2,709	1,001	227	458	452	0.87	4.2	9	14	18	32	43
	S4	1,042	829	1,037	148	214	210	0.93	3.8	8	12	13	20	23
	S5	4,399	3,296	977	191	359	352	0.9	4.2	11	16	21	37	46
	S6	1,240	937	1,039	199	318	313	0.89	4.2	9	12	14	26	31
ALPS	S1	5,958	2,097	1,017	78	100	97	0.99	3.2	5	8	9	15	18
	S2	12,619	3,168	998	168	209	290	0.93	3.8	6	10	14	24	32
	S3	3,730	1,276	1,027	139	217	212	0.95	3.8	4	6	7	18	21
	S4	5,025	2,097	995	123	181	177	0.98	3.1	5	8	12	22	27
	S5	9,482	3,110	1,012	107	143	140	0.97	3.4	4	6	9	14	18
	S6	9,854	2,038	999	195	238	237	0.98	4	5	7	12	23	29
P value (Student's *t* test)		<0.05*	<0.05*	NS	<0.1*	<0.05*	<0.05*	<0.01*	<0.05*	<0.01*	<0.01*	<0.01*	<0.01*	<0.01*

[a] Number of filtered reads (after initial processing), number of unique reads, and number of reads after normalization per library, species richness indices (S_{obs}, S_{est}, and Chao1), Good's coverage, alpha diversity (Shannon-Wiener index, H′), and taxonomy (numbers of phyla, classes, orders, families, and genera). *, statistically significant result; NS, nonsignificant.

ment was investigated by analysis of 3,011 available metagenome data sets in the IMG/M database. The *phoD* gene was found in a range of environments (Fig. 1b), with the greatest abundance in soil, followed by marine and air environments.

Metagenomic studies focusing on phosphatases in marine environments have shown that *phoD* and *phoX* are more common than *phoA* in these samples (8, 38). The high diversity and relative abundance of the *phoD* gene found in soil metagenomes (Fig. 1b) suggest that *phoD* may also be particularly relevant in terrestrial ecosystems, although the relative abundances of the three alkaline phosphatase families in soil have not yet been studied on the metagenome level. The fact that organic P represents between 30% and 80% of the total P in grassland and agricultural soils, mainly in the form of diverse phosphomonoesters and diesters (3), may promote the diversity of *phoD* in terrestrial ecosystems.

Performance of PHOD and ALPS primers. A key aim of this work was to design a new set of PHOD primers targeting the bacterial *phoD* alkaline phosphatase for studying the *phoD* bacterial community diversity and composition in soil. The PHOD primers were tested on six soils that represent a range of contrasting soil properties, collected in Australia and Switzerland, and the results were compared with those obtained with the same samples using the ALPS primers.

Generally, amplification using PHOD primers resulted in fewer filtered reads than that with ALPS primers, with 2,309 ± 1,148 (mean ± standard deviation) and 7,778 ± 3,107 reads and average read lengths of 380 ± 33 bp and 364 ± 35 bp for PHOD- and ALPS-amplified samples, respectively (Table 2). The difference in the number of filtered reads per library was directly linked to primer design, more particularly to the degree of degeneracy of the PHOD primers. Increasing degeneracy in primers generally reduces PCR efficiency due to the dilution of each unique primer sequence (39). Degenerate primers increase the risk of unspecific annealing during the PCR but increase the probability of amplifying yet-unknown *phoD* gene sequences by allowing all coding possibilities for an amino acid residue in the nucleic acid sequences (40). When used appropriately, degenerate primers, such as the PHOD primers, represent a great advantage in studies on genetic diversity by targeting known and unknown sequences in environmental samples (41).

By filtering out redundant sequences, the number of reads decreased remarkably in the ALPS-amplified samples, leading to more-similar numbers of unique reads for the two sets of primers, which averaged 1,893 ± 885 bp (mean ± standard deviation) and 2,297 ± 659 bp for PHOD- and ALPS-amplified samples, respectively. This showed that ALPS-amplified samples consisted of a greater number of redundant reads than did PHOD-amplified samples. Finally, normalization of the library size in order to compare the two primer sets resulted in an average library size of 1,013 ± 31 bp. Our results suggest that the ALPS primers target a narrow spectrum of sequences which represent a large fraction of the reads after amplification.

Species richness and alpha diversity of the *phoD* gene in six soils. Amplification with PHOD primers revealed a 2-fold variation in species richness among the six samples (Table 2). S_{obs} was lowest in S4 and highest in S1, with 148 and 290 OTUs, respectively. Chao1 and S_{est} indices, derived from the rarefaction curves, showed a similar trend. The difference in species richness between samples is well illustrated by the rarefaction curves (Fig. 2a). The rarefaction curve of S1 had the steepest slope, showing the greatest increase of species with the number of reads, while that of S4 reached the asymptote with the fewest reads (ca. 3,000).

Compared to amplification with PHOD primers, amplification with ALPS primers resulted in significantly lower species richness and alpha diversity (Table 2). In ALPS-amplified samples, the rarefaction curves always reached the asymptote with fewer reads than in the corresponding PHOD-amplified samples (Fig. 2a and b). The rarefaction curves of S1 when amplified using PHOD and ALPS primers contrasted particularly strongly, leading to a 7-fold difference in Chao1 and S_{est}. Likewise, H′ was always greater in PHOD- than in ALPS-amplified samples. This

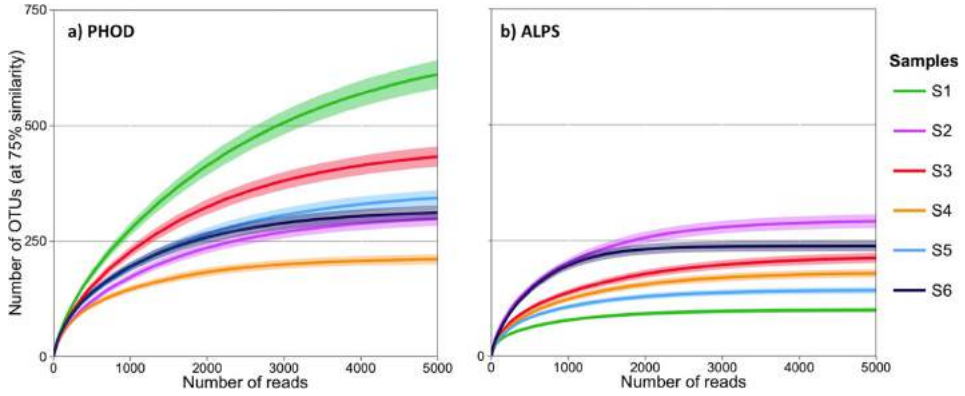


FIG 2 Rarefaction curves of samples S1 to S6 amplified by PHOD (a) and ALPS (b) primers extrapolated to 5,000 reads with 95% confidence intervals.

suggests that PHOD primers target a broader diversity of *phoD*-bearing bacteria than ALPS primers.

Using ALPS primers, Tan et al. (10) found between 450 and 548 OTUs in soils fertilized with zero, medium, or high P input, with a sequencing depth between 14,279 and 16,140 reads. In contrast, Fraser et al. (16) reported lower numbers, which are in the same range as in the six soils analyzed in this study. They found between 137 and 163 OTUs in soils from organic and conventional cropping systems and prairie, with a sequencing depth of 11,537 to 54,468 reads. Thus, the number of OTUs seems to be quite variable between studies and/or soils. By applying both primers on the same soils, we found that PHOD primers targeted a larger species spectrum than ALPS primers.

Dominant phyla harboring *phoD* in six soils. Taxonomy was assigned to most sequences using BLAST+ (27) (Fig. 3; see also Table S2 in the supplemental material). The remainder, 5,052 reads representing between 0.1 and 22% of the total filtered read number, were not assigned a taxonomic identity. In theory, the primers could amplify *phoD* in archaea and eukaryotes also, as *phoD* has been found in several archaeal and eukaryotic species in the IMG/M database. In the six soils studied here, both ALPS and PHOD primers amplified *phoD* from bacteria only, based on identification using BLAST+.

PHOD primers targeted *phoD* genes from 13 phyla (*Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Nitrospirae*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, and *Verrucomicrobia*). They covered 22 classes, 38 orders, 71 families, and 113 genera. The dominant orders were *Actinomycetales* (13 to 35%), *Bacillales* (1 to 29%), *Gloeobacterales* (1 to 18%), *Rhizobiales* (18 to

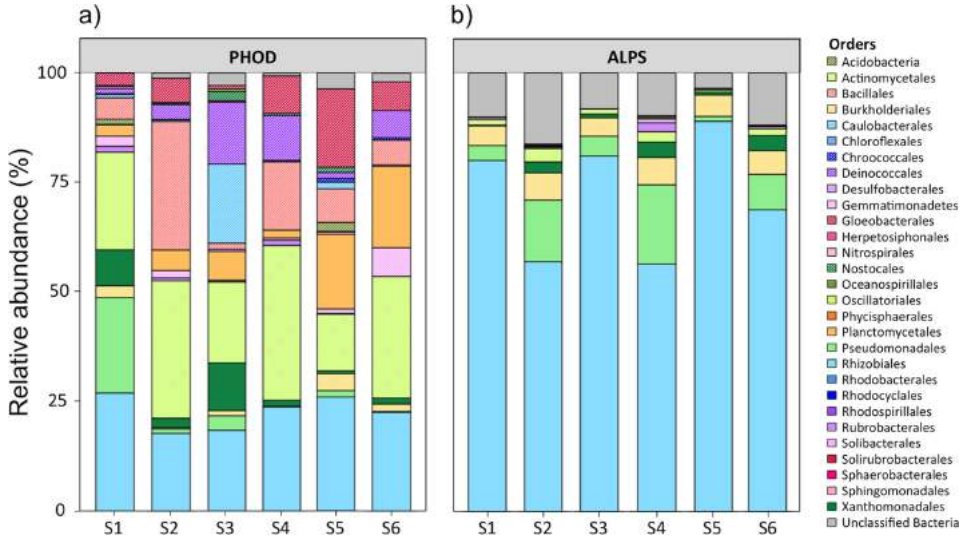


FIG 3 Relative abundance as a percentage of the total community at the order level in samples S1 to S6 amplified by PHOD (a) and ALPS (b) primers.

27%), and *Pseudomonadales* (0 to 22%). A libshuff analysis showed that the *phoD* bacterial communities in the different soils were significantly different from each other ($P < 0.001$). S1 was characterized by 25% *Pseudomonadales* and 10% *Xanthomonadales*. The highest relative abundance of *Bacillales* (29%) was found in S2. S3 had particularly high abundances of *Caulobacterales* (19%), *Deinococcales* (14%), and *Xanthomonadales* (11%). *Planctomycetes* were especially abundant in S4 and S6, with 18 and 19%, respectively, while S5 showed a high abundance of *Gloeobacterales* (18%).

ALPS primers amplified *phoD* genes from 6 phyla (*Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Cyanobacteria*, *Gemmatimonadetes*, and *Proteobacteria*). In more detail, ALPS primers covered 13 classes, 22 orders, 42 families, and 64 genera. The most prevalent class was *Alphaproteobacteria* (55 to 92%). *Rhizobiales* was the dominant taxon in this class, with an overrepresentation of *Methylobacterium* sp., which represented 60 to 95% of the abundance of *Rhizobiales*. A libshuff analysis showed that the structures of the *phoD* bacterial communities in the different samples were also significantly different from each other ($P < 0.001$).

This taxonomy analysis highlights the fact that the *phoD* gene is widespread across phyla and that the PHOD primers covered the *phoD* diversity well. PHOD primers targeted *phoD* genes in 13 out of the 20 phyla known to carry the *phoD* gene, based on the IMG/M database. PHOD primers captured a particularly large diversity of *Actinobacteria*, including the common soil genera *Actinomyces*, *Arthrobacter*, *Kineococcus*, *Kitasatospora*, *Micrococcus*, and *Streptosporangium* (42), and of *Proteobacteria*, including *Azorhizobium*, *Rhodospirillum*, *Caulobacter*, *Geobacter*, and *Variovorax* (43). Both *Actinobacteria* and *Proteobacteria* are known to be important for mineralization of soil organic matter and in composting processes (44, 45). Our sequencing results for soils, in accordance with the IMG/M analysis, show that a greater diversity of microorganisms than previously thought contributes to organic P mineralization by secreting PhoD.

PHOD primers amplified many sequences belonging to phyla with low abundances in the IMG/M database. These sequences were affiliated with the phyla *Deinococcus-Thermus* (e.g., *Deinobacter* sp.), *Nitrospirae* (e.g., *Nitrospira* sp.), *Spirochaetes* (e.g., *Spirochaeta* sp.), *Planctomycetes* (e.g., *Isosphaera* sp. and *Planctomyces* sp.), and *Verrucomicrobia* (e.g., *Opitutus* sp.). The ALPS primers did not amplify *phoD* genes from these phyla. Moreover, compared to the PHOD primers, the ALPS primers failed to detect *phoD* genes from many genera, including, e.g., *Anabaena*, *Chroococcidiopsis*, and *Chroococcus*, belonging to the *Cyanobacteria*. Our results support the conclusion of Tan et al. (10) that the ALPS primers have an amplification bias, restraining the amplification to a limited number of microbial taxa and overrepresenting *Alphaproteobacteria*, probably because of the few sequences used to design the primers (7 sequences from 4 phyla used, compared with 315 sequences from 11 phyla used here for the primer design).

Soil pH is the main driver of the *phoD* bacterial community. Redundancy analysis (RDA) of the PHOD-amplified samples indicated that 49.1% of the variation was explained by the two main RDA components (Fig. 4). Variance partitioning analysis showed that soil pH explained 23.7% and total P 18.3% of the variance among the communities. However, soil pH was the only environmental variable that was significantly correlated with the distribution of the samples ($P = 0.03$). The most divergent samples along

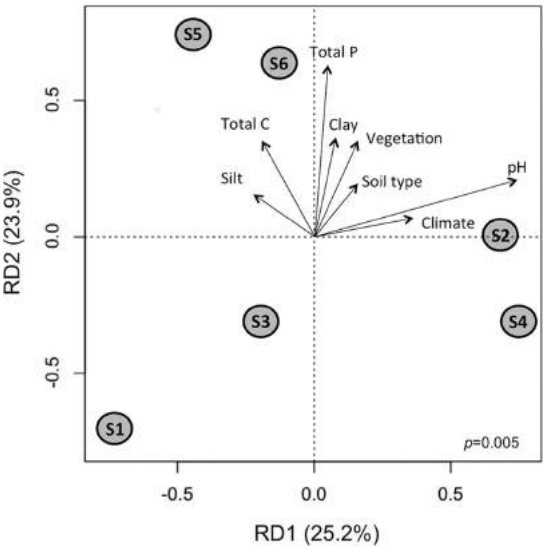


FIG 4 Redundancy analysis of the *phoD* bacterial community of samples S1 to S6 amplified by PHOD primers with the environmental variables clay and silt content, total C and P, soil type, climate, vegetation and soil pH. The significance of the model is indicated in the bottom right corner. Note that soil pH was the unique environmental variable that was significantly correlated with the *phoD* bacterial community ($P = 0.03$).

the first RDA component axis were S1 and S4. The observed differences between these samples are likely due to the very contrasting soil and environmental properties between the sampling sites. S1 was taken from an oceanic and temperate climatic region with dense vegetation, while S4 was collected in a hot semiarid climatic region with only scattered vegetation, where lower soil microbial biomass and diversity are expected (46). S1 and S4 also exhibited the biggest difference in soil pH, which is regarded as the main environmental driving force that affects total microbial communities and activities (47, 48). Soil pH has previously been observed to be an important driver of the *phoD* bacterial community in the rhizosphere of wheat grown in different soils (15). Phosphatase activity can respond to changes in soil pH within days, e.g., after a lime treatment in agricultural soils (49). The second RDA component was linked mainly to total P. The *phoD* communities of S5 and S6 clustered together along the second-component axis, probably because these two samples were both collected in Switzerland and had high total carbon and other similar soil properties. In contrast, S1, S3, and S4 had low total C and P values.

Previous studies using the ALPS primers have reported an effect of the application of organic and conventional fertilizers, crop management, vegetation, and pH (10, 12–16, 50). The plant community has been reported to have an impact on *phoD* diversity and community structure in monocultures (14, 15). P fertilization has been reported to either increase (10) or reduce (12) the diversity of the *phoD* gene. Jorquera et al. (13) observed that P fertilization alone did not affect the *phoD* bacterial community structure in a Chilean Andisol pasture, while combined N and P fertilization did change the *phoD* bacterial community structure. While all these studies have provided some insights into the environmental driv-

ers affecting *phoD* bacterial communities, they need to be interpreted with caution due to the amplification bias of the ALPS primers toward *Alphaproteobacteria* described above. PHOD primers should now be applied to a wider range of soils to verify whether pH is the main driver of the *phoD* bacterial community.

In conclusion, evaluation of metagenomic data sets revealed that the *phoD* gene is found primarily in bacteria and is spread across 20 bacterial phyla. *phoD* has been found to be ubiquitous in the environment, with terrestrial ecosystem metagenomes containing the highest relative abundance of *phoD*. The newly designed PHOD primers reported here covered the large diversity of the *phoD* gene better than previously published primers and amplified sequences affiliated with 13 bacterial phyla. The most prevalent *phoD* genes identified in six diverse soils from Europe and Australia were affiliated with the orders *Actinomycetales*, *Bacillales*, *Gloeobacterales*, *Rhizobiales*, and *Pseudomonadales*. Soil pH was found to be the main environmental driver affecting the *phoD* bacterial community. PHOD primers can be used as a tool to study *phoD* bacterial community diversity and composition and to identify and quantify microorganisms that carry and express *phoD* in the environment.

ACKNOWLEDGMENTS

We thank Stefan Zoller for the Perl scripts for the taxonomic analysis and the Genetic Diversity Center (Zurich, Switzerland) for technical assistance. We also acknowledge Agroscope (Switzerland) and the New South Wales Department of Primary Industry (NSW, Australia) for access to the sampling sites.

This work was supported by the Swiss National Science Foundation (SNF) and by a research grant from the University of Sydney.

We declare no conflicts of interest.

REFERENCES

1. **Westheimer FH.** 1987. Why nature chose phosphates. Science **235:**1173–1178. http://dx.doi.org/10.1126/science.2434996.
2. **Vitousek PM, Porder S, Houlton BZ, Chadwick OA.** 2010. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions. Ecol Appl **20:**5–15. http://dx.doi.org/10.1890/08-0127.1.
3. **Condron LM, Turner BL, Cade-Menun BJ.** 2005. Chemistry and dynamics of soil organic phosphorus, p 87–122. *In* Sims JT, Sharpley AN (ed), Phosphorus: agriculture and the environment. ASA, CSSA and SSSA, Madison, WI.
4. **Vershinina OA, Znamenskaya LV.** 2002. The Pho regulons of bacteria. Microbiology **71:**497–511. http://dx.doi.org/10.1023/A:1020547616096.
5. **Kageyama H, Tripathi K, Rai AK, Cha-um S, Waditee-Sirisattha R, Takabe T.** 2011. An alkaline phosphatase/phosphodiesterase, PhoD, induced by salt stress and secreted out of the cells of *Aphanothece halophytica*, a halotolerant cyanobacterium. Appl Environ Microbiol **77:**5178–5183. http://dx.doi.org/10.1128/AEM.00667-11.
6. **Wu JR, Shien JH, Shieh HK, Hu CC, Gong SR, Chen LY, Chang PC.** 2007. Cloning of the gene and characterization of the enzymatic properties of the monomeric alkaline phosphatase (PhoX) from *Pasteurella multocida* strain X-73. FEMS Microbiol Lett **267:**113–120. http://dx.doi.org/10.1111/j.1574-6968.2006.00542.x.
7. **Boulanger RR, Kantrowitz ER.** 2003. Characterization of a monomeric *Escherichia coli* alkaline phosphatase formed upon a single amino acid substitution. J Biol Chem **278:**23497–23501. http://dx.doi.org/10.1074/jbc.M301105200.
8. **Luo H, Benner R, Long RA, Hu J.** 2009. Subcellular localization of marine bacterial alkaline phosphatases. Proc Natl Acad Sci U S A **106:**21219–21223. http://dx.doi.org/10.1073/pnas.0907586106.
9. **Zaheer R, Morton R, Proudfoot M, Yakunin A, Finan TM.** 2009. Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. Environ Microbiol **11:**1572–1587. http://dx.doi.org/10.1111/j.1462-2920.2009.01885.x.
10. **Tan H, Barret M, Mooij MJ, Rice O, Morrissey JP, Dobson A, Griffiths B, O'Gara F.** 2013. Long-term phosphorus fertilisation increased the diversity of the total bacterial community and the *phoD* phosphorus mineraliser group in pasture soils. Biol Fertil Soils **49:**661–672. http://dx.doi.org/10.1007/s00374-012-0755-5.
11. **Sakurai M, Wasaki J, Tomizawa Y, Shinano T, Osaki M.** 2008. Analysis of bacterial communities on alkaline phosphatase genes in soil supplied with organic matter. Soil Sci Plant Nutr **54:**62–71. http://dx.doi.org/10.1111/j.1747-0765.2007.00210.x.
12. **Chhabra S, Brazil D, Morrissey J, Burke J, O'Gara F, Dowling DN.** 2013. Fertilization management affects the alkaline phosphatase bacterial community in barley rhizosphere soil. Biol Fertil Soils **49:**31–39. http://dx.doi.org/10.1007/s00374-012-0693-2.
13. **Jorquera MA, Martínez OA, Marileo LG, Acuña JJ, Saggar S, Mora ML.** 2014. Effect of nitrogen and phosphorus fertilization on the composition of rhizobacterial communities of two Chilean Andisol pastures. World J Microbiol Biotechnol **30:**99–107. http://dx.doi.org/10.1007/s11274-013-1427-9.
14. **Wang Y, Marschner P, Zhang F.** 2012. Phosphorus pools and other soil properties in the rhizosphere of wheat and legumes growing in three soils in monoculture or as a mixture of wheat and legume. Plant Soil **354:**283–298. http://dx.doi.org/10.1007/s11104-011-1065-7.
15. **Wang Y, Zhang F, Marschner P.** 2012. Soil pH is the main factor influencing growth and rhizosphere properties of wheat following different pre-crops. Plant Soil **360:**271–286. http://dx.doi.org/10.1007/s11104-012-1236-1.
16. **Fraser TD, Lynch DH, Bent E, Entz MH, Dunfield KE.** 2015. Soil bacterial *phoD* gene abundance and expression in response to applied phosphorus and long-term management. Soil Biol Biochem **88:**137–147. http://dx.doi.org/10.1016/j.soilbio.2015.04.014.
17. **Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M.** 2012. IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res **40:**D123–D129. http://dx.doi.org/10.1093/nar/gkr975.
18. **Huang Y, Niu B, Gao Y, Fu L, Li W.** 2010. CD-HIT suite: a web server for clustering and comparing biological sequences. Bioinformatics **26:**680–682. http://dx.doi.org/10.1093/bioinformatics/btq003.
19. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792–1797. http://dx.doi.org/10.1093/nar/gkh340.
20. **Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R.** 2011. PrimerProspector: de novo design and taxonomic analysis of barcoded PCR primers. Bioinformatics **27:**1159–1161. http://dx.doi.org/10.1093/bioinformatics/btr087.
21. **Gulvik CA, Effler CT, Wilhelm SW, Buchen A.** 2012. De-MetaST-BLAST: a tool for the validation of degenerate primer sets and data mining of publicly available metagenomes. PLoS One **7:**e50362. http://dx.doi.org/10.1371/journal.pone.0050362.
22. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J Mol Biol **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
23. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:**2194–2200. http://dx.doi.org/10.1093/bioinformatics/btr381.
24. **Rodriguez F, Lillington J, Johnson S, Timmel CR, Lea SM, Berks BC.** 2014. Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. J Biol Chem **289:**30889–30899. http://dx.doi.org/10.1074/jbc.M114.604892.
25. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75:**7537–7541. http://dx.doi.org/10.1128/AEM.01541-09.
26. **Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D.** 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol **9:**e1000602. http://dx.doi.org/10.1371/journal.pbio.1000602.
27. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10:**421. http://dx.doi.org/10.1186/1471-2105-10-421.
28. **Chao AC, Shen T-J.** 2003. Nonparametric estimation of Shannon's index

of diversity when there are unseen species in sample. Environ Ecol Stat **10:**429–443. http://dx.doi.org/10.1023/A:1026096204727.
29. **Good IJ.** 1953. The population frequencies of species and the estimation of population parameters. Biometrika **40:**237–264. http://dx.doi.org/10.1093/biomet/40.3-4.237.
30. **Gotelli NJ, Colwell RK.** 2011. Estimating species richness, p 39–54. *In* Magurran AE, McGill BJ (ed), Biological diversity: frontiers in measurement and assessment, vol 12. Oxford University Press, Oxford, United Kingdom.
31. **Schloss PD, Larget BR, Handelsman J.** 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl Environ Microbiol **70:**5485–5492. http://dx.doi.org/10.1128/AEM.70.9.5485-5492.2004.
32. **Sait M, Hugenholtz P, Janssen PH.** 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environ Microbiol **4:**654–666. http://dx.doi.org/10.1046/j.1462-2920.2002.00352.x.
33. **Zappa S, Rolland J-L, Flament D, Gueguen Y, Boudrant J, Dietrich J.** 2001. Characterization of a highly thermostable alkaline phosphatase from the euryarchaeon *Pyrococcus abyssi*. Appl Environ Microbiol **67:**4504–4511. http://dx.doi.org/10.1128/AEM.67.10.4504-4511.2001.
34. **Wende A, Johansson P, Vollrath R, Dyall-Smith M, Oesterhelt D, Grininger M.** 2010. Structural and biochemical characterization of a halophilic archaeal alkaline phosphatase. J Mol Biol **400:**52–62. http://dx.doi.org/10.1016/j.jmb.2010.04.057.
35. **Šnajdr J, Valšková V, Merhautová V, Cajthaml T, Baldrian P.** 2008. Activity and spatial distribution of lignocellulose-degrading enzymes during forest soil colonization by saprotrophic basidiomycetes. Enzyme Microb Technol **43:**186–192. http://dx.doi.org/10.1016/j.enzmictec.2007.11.008.
36. **Dyhrman ST, Ruttenberg KC.** 2006. Presence and regulation of alkaline phosphatase activity in eukaryotic phytoplankton from the coastal ocean: implications for dissolved organic phosphorus remineralization. Limnol Oceanogr **51:**1381–1390. http://dx.doi.org/10.4319/lo.2006.51.3.1381.
37. **Fernandez NJ, Kidney BA.** 2007. Alkaline phosphatase: beyond the liver. Vet Clin Pathol **36:**223–233. http://dx.doi.org/10.1111/j.1939-165X.2007.tb00216.x.
38. **Sebastian M, Ammerman JW.** 2009. The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. ISME J **3:**563–572. http://dx.doi.org/10.1038/ismej.2009.10.
39. **Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF.** 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol **71:**8966–8969. http://dx.doi.org/10.1128/AEM.71.12.8966-8969.2005.
40. **Limansky AS, Viale AM.** 2002. Can composition and structural features of oligonucleotides contribute to their wide-scale applicability as random PCR primers in mapping bacterial genome diversity? J Microbiol Methods **50:**291–297. http://dx.doi.org/10.1016/S0167-7012(02)00040-4.
41. **Menzel P, Stadler PF, Gorodkin J.** 2011. maxAlike: maximum likelihood-based sequence reconstruction with application to improved primer design for unknown sequences. Bioinformatics **27:**317–325. http://dx.doi.org/10.1093/bioinformatics/btq651.
42. **Bora N, Ward A.** 2008. The actinobacteria. *In* Goldman E, Green L (ed), Practical handbook of microbiology. CRC Press, Boca Raton, FL.
43. **Nacke H, Thürmer A, Wolher A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Rolf D.** 2011. Pyrosequencing-based assessment of bacterial community structure alon different management types in German forest and grassland soils. PLoS One **6:**e17000. http://dx.doi.org/10.1371/journal.pone.0017000.
44. **Danon M, Franke-Whittle IH, Insam H, Chen Y, Hadar Y.** 2008. Molecular analysis of bacterial community succession during prolonged compost curing. FEMS Microbiol Ecol **65:**133–144. http://dx.doi.org/10.1111/j.1574-6941.2008.00506.x.
45. **Yu H, Zeng G, Huang H, Xi X, Wang R, Huang D, Huang G, Li J.** 2007. Microbial community succession and lignocellulose degradation during agricultural waste composting. Biodegradation **18:**793–802. http://dx.doi.org/10.1007/s10532-007-9108-8.
46. **Bachar A, Al-Ashhab A, Soares MIM, Sklarz MY, Angel R, Ungar ED, Gillor O.** 2010. Soil microbial abundance and diversity along a low precipitation gradient. Microb Ecol **60:**453–461. http://dx.doi.org/10.1007/s00248-010-9727-1.
47. **Lauber CL, Hamady M, Knight R, 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. http://dx.doi.org/10.1128/AEM.00335-09.
48. **Fierer N, Jackson RB.** 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103:**626–631. http://dx.doi.org/10.1073/pnas.0507535103.
49. **Dick WA, Cheng L, Wang P.** 2000. Soil acid and alkaline phosphatase activity as pH adjustment indicators. Soil Biol Biochem **32:**1915–1919. http://dx.doi.org/10.1016/S0038-0717(00)00166-8.
50. **Fraser T, Lynch DH, Entz MH, Dunfield KE.** 2014. Linking alkaline phosphatase activity with bacterial *phoD* gene abundance in soil from a long-term management trial. Geoderma **257–258:**115–122. http://dx.doi.org/10.1016/j.geoderma.2014.10.016.
51. **Kottek M, Grieser J, Beck C, Rudolf B, Rubel F.** 2006. World map of the Köppen-Geiger climate classification updated. Meteorol Z **15:**259–263. http://dx.doi.org/10.1127/0941-2948/2006/0130.
52. **Food and Agriculture Organization of the United Nations/International Soil Reference and Information Centre/International Society of Soil Science.** 1998. World reference base for soil resources. Food and Agriculture Organization of the United Nations, Rome, Italy.
53. **Anderson JM, Ingram JSI.** 1993. Tropical soil biology and fertility: a handbook of methods, 2nd ed. CAB International, Wallingford, Oxon, United Kingdom.
54. **Ohno R, Zibilske LM.** 1991. Determination of low concentrations of phosphorus in soil extracts using malachite green. Soil Sci Soc Am J **55:**892–895. http://dx.doi.org/10.2136/sssaj1991.03615995005500030046x.

A.2 Taxonomic and environmental prevalence of the *acpA* acid phosphatase gene - a meta-analysis

A.2.1 Introduction

Acid phosphatases are produced by prokaryotes (e.g. archaea, bacteria) and eukaryotes (e.g. fungi, plants) (Richardson et al. 2000; Richardson and Simpson 2011). They are encoded by a large variety of genes (Table 1.1) such as the *acpA* gene. AcpA is a dimeric acid phosphatase that requires a metal ion as co-factor, although the latter has not yet been identified (Felts et al. 2006). The core domain is a twisted 8-stranded β-sheet flanked by three α-helices on either side, with the active site located above the carboxyl-terminal edge of the β-sheet (Felts et al. 2006). The protein structure resembles that of PhoD and PhoX alkaline phosphatase (Felts et al. 2006). AcpA is defined as a non-specific acid phosphatase and has an optimal activity at pH 6 to 7 (Costas et al. 2010; Reilly et al. 1996).

The *acpA* gene is principally found in bacteria, but has also been found in protists (e.g. Amoebozoa). The *acpA* gene has been found in multiple bacteria including *Beta-* (e.g. *Burkholderia* spp.) and *Gamma-Proteobacteria* (e.g. *Pseudomonas* fluorescens and *Francisella* tularensis) and *Actinobacteria* (e.g. *Mycobacteria* sp.). It is widely spread in *Burkholderia* species such as *B. pseudomallei*, *B. thalandensis* and *B. mallei* (Burtnick et al. 2001). However, the key microorganisms that harbour the *acpA* gene and the prevalence of the *acpA* in the environment are unknown.

A.2.2 Material and Methods

The prevalence of the *acpA* gene was assessed using the Integrated Microbial Genomes and Metagenomes (IMG/M) database (Markowitz et al. 2012). Draft and complete

genome datasets were used to evaluate the distribution of *acpA* across archaeal, bacterial and fungal phyla, and metagenome datasets were used to evaluate the prevalence of *acpA* in the environment (data accessed on November 7th 2015). Metagenome datasets were normalized as number of *acpA* counts per number of bases sequenced per metagenome dataset and categorized after the classification of the IMG/M database.

A.2.3 Results and Discussion

A.2.3.1 Taxonomic distribution of the *acpA* alkaline phosphatase gene

Our current knowledge of the taxonomic distribution of the *acpA* gene was evaluated using the IMG/M database. Among all bacterial phyla in the database, the *acpA* gene was present in 16% of the genomes. The *acpA* gene was reported in 679 bacterial genomes affiliated to 8 bacterial phyla (Figure A.1). In the *Proteobacteria*, *Beta-Proteobacteria*, the most frequently isolated and sequenced bacterial order, represented 59% of the genome containing at least one *acpA* homologue, followed by *Alpha-Proteobacteria* (12%) and *Gamma-Proteobacteria* (29%). *Burkholderia* sp. represented 356 of the 679 genomes in the database. The number of *acpA* homologue copies per genome varied between 1 and 6, with 61% containing only 1 copy.

acpA has been found in several bacteria including *Burkholderia* sp., *Francisella* sp., *Pseudomonas* sp., *Mycobacter* sp., and *Bordetella* sp. In the literature, *acpA* has been principally studied in pathogenic bacteria such as *Burkholderia* spp. Species of *Burkholderia* spp. groups can be free-living or pathogenic bacteria. Free-living *Burkholderia* spp. are commonly found in aquatic and terrestrial ecosystems, particularly in the rhizosphere, as free-living organisms (Vial et al. 2011; Zhang and Xie 2007). On the other hand, *Burkholderia* spp. also represent a life-threat for patients with cystic fibrosis and are the etiologic agents of melioidosis. Pathogenic *Burkholderia* spp. are intensively studied and this could explain their large presence in the genomic database.

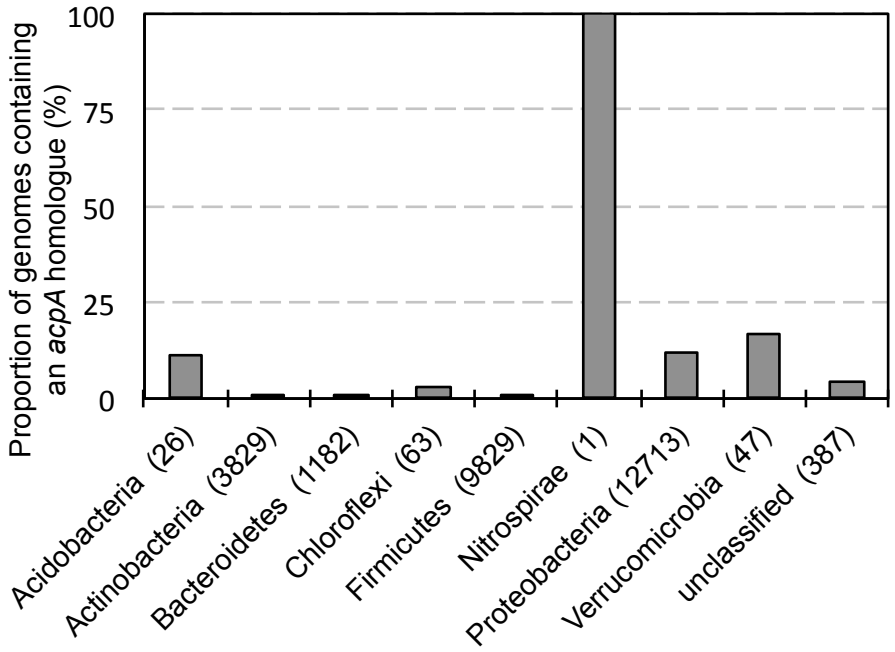


Figure A.1. Proportion of sequenced genomes containing a *acpA* homologue in the IMG/M database (on September 10th 2015). Numbers in brackets indicate the total number of sequenced genomes in each phylum.

Compared to *phoD* and *phoX*, *acpA* is not as widely spread across bacterial phyla according to the IMG/M database. Only a small fraction harbor the *acpA* gene. The number of different acid phosphatase genes in databases such as NCBI and IMG/M suggest that the variety of acid phosphatase genes is remarkably larger than that of alkaline phosphatase genes. Hence, bacteria may harbor multiple acid phosphatase genes in addition to *acpA* .

A.2.3.2 Environmental prevalence of the *acpA* alkaline phosphatase gene

Among the 4,326 metagenomes of the IMG/M database, 677 metagenomes bore at least one *acpA* homologue. Soil metagenomes represented the type of environments most

frequently carrying *acpA* homologues, followed by arthropod-associated microbiome and fresh water environments (Figure A.2).

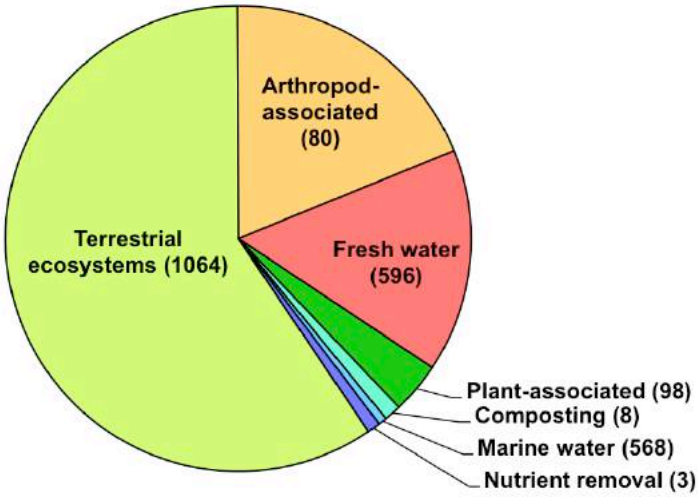


Figure A.2. Relative abundance of *acpA* genes in different types of environments (normalized as number of *acpA* counts per number of bases sequenced per metagenome dataset). Numbers in brackets indicate the number of metagenome datasets per environment type.

The *acpA* gene has mainly been found in pathogenic bacteria isolated from animals such as *Burkholderia pseudomallei* and *Pseudomonas aeruginosa* . Known to be opportunistic pathogens infecting patients with cystic fibrosis, these bacterial species are also commonly found in soil as free-living microorganisms (Zhang and Xie 2007). The *acpA* gene is found in the pathogenic bacteria, *Francisella tularensis* , which is causing tulameria, a skin infectious disease. The tulameria is commonly transmitted via arthropod bite (Dennis et al. 2001). Studies on vectors of the tulameria suggest that *Francisella tularensis* is part of the natural flora of certain arthropods, and thus, also present in water ecosystems (Dennis et al. 2001).

A.2.4 Conclusion

This is the first study summarizing the taxonomic distribution and environmental prevalence of the *acpA* gene. *acpA* is found mainly in *Proteobacteria* and largely found in terrestrial environments.

A.3 Design and validation of *acpA*-targeting primers

A.3.1 Introduction

The AcpA acid phosphatase is principally described as a phosphomonoesterase with a nonspecific substrate spectrum. In addition to its phosphomonoesterase activity, it also has phosphodiesterase, phospholipase and pyrophosphatase activity (Costas et al. 2010; Reilly et al. 1996). Its optimum enzymatic activity is at pH 6 to 7.2 (Reilly et al. 1996; Stonehouse et al. 2002). The structure of AcpA includes an 8-stranded Beta sheet flanked by 3 α-helices on each side and requires cations such as Ca^{2+} or Mg^{2+} as co-factor (Felts et al. 2006). The structure of AcpA resembles that of the PhoD and PhoX alkaline phosphatases (Felts et al. 2006). Multiple copies of *acpA* homologues can be found in the same genome (Stonehouse et al. 2002). The *acpA* gene is part of the Pho regulon regulated by the P availability in the environment (Stonehouse et al. 2002) and is about 1,734 bp long (Costas et al. 2010; Reilly et al. 1996). The enzyme is synthesized in the cytoplasm and is exported to the periplasm via the TAT transportation pathway (Stonehouse et al. 2002; Felts et al. 2006). However, the key microorganisms that harbour the *acpA* gene and the prevalence of the *acpA* in the environment are unknown. Here, we designed a new set of primers to target the *acpA* gene in soil microorganisms and tested them on the soil samples of Chapter 4.

A.3.2 Material and Methods

A.3.2.1 Soil sampling

Five sites were sampled in Australia in spring 2013 and five sites in Switzerland in summer 2014 (Table A.1), covering a broad range of soil types, vegetation and climatic conditions varying from hot semi-arid to continental temperate climates. At each site,

soil from three land-use types (arable, forest and grassland soils), was collected. For each sampling, five soil cores from the upper 5 cm were randomly collected and homogenized by sieving (4 mm). A subsample was stored at -20°C for molecular analysis.

A.3.2.2 DNA extraction from soil

Genomic DNA was extracted from 0.25 g frozen soil using the PowerSoil® DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) following the manufacturer's instructions with a slight modification in the homogenization and cell lysis step: TissueLyzer II (Qiagen, Valencia, CA, USA) was used for 2 times 3 min at 30 Hz. DNA was eluted in 45 µL elution buffer and stored at -20°C until further processing. DNA concentration and quality were assessed with a micro-volume UV/VIS spectrophotometer (UVS-99, UVISDROP; ACTGene, Inc., Piscataway, NJ, USA).

A.3.2.3 Design and validation of the *acpA* acid phosphatase gene targeting primers

Gene sequences annotated as *acpA* and/or associated with TIGR03397 (Lee et al. 2005) (TIGR Gene Index database; http://compbio.dfci.harvard.edu/tgi/), which corresponds to *acpA* acid phosphatase, were retrieved from the European Nucleotide Archive (ENA) and Uniprot Knowledgebase (UniprotKB) databases. They were then clustered at 97% similarity using CD-HIT (Huang et al. 2010), resulting in a total of 45 sequences used as reference for the primer design (see list in Table A.1). Sequences were affiliated to 3 phyla including *Euarcheota* (4 sequences), *Actinobacteria* (5 sequences), *Proteobacteria* (1, 25 and 10 *Alpha-*, *Beta-* and *Gamma-Proteobacteria* sequences, respectively).

Alignment was constructed using MUSCLE (Edgar 2004) and manually reviewed and improved in Geneious 6.1.2 (Biomatters, Australia, http://www.geneious.com). The most suitable regions for primer design were identified using PrimerProspector (Walters et al. 2011). Forward and reverse candidate primers were then manually designed to reach the maximum coverage of the reference sequences. Candidate primer pairs were tested

in-silico using De-metaST-BLAST (Gulvik et al. 2012) to identify potential primer pairs with an appropriate product size and coverage of the reference sequences.

Candidate primers (6 forward and 4 reverse primers) were tested in a gradient PCR using a mixture of soil genomic DNA containing soil from AUS3-A and CH4-G as template (See Table 4.1). PCR reactions were performed in a 25 μL volume containing MyTaq Reaction Buffer (including $MgCl_2$ and dNTPs), 0.5 μM of each primer and 0.6 Units of MyTaq polymerase (Bioline, NSW, Australia) with 1-2 ng DNA as template in an S1000 thermocycler (Bio-Rad Laboratories, CA). The amplification reaction included an initial denaturation step of 5 min at 95°C followed by 35 cycles of a denaturation step of 30 s at 95°C, an annealing step of 30 s at the calculated optimal annealing temperature of each candidate primer pair (gradient of ± 3 degrees) and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. Amplicon size and intensity, and the presence of primer dimers, was assessed visually by electrophoresis on a 1.5% (w/v) agarose gel stained with ethidium bromide.

Amplicon specificity was evaluated for selected primer pairs by cloning and sequencing. PCR products were ligated at 4 °C overnight using pGEM®-T Vector Systems (Promega, Madison, WI) and transformed into chemically competent *E. coli* cells (α-select; F-*(*deo)R *end*A1 *rec*A1 *rel*A1 *gyr*A96 *hsd*R17 (rk^-, mk^+) E44 thi-1 *pho*A Δ(*lac*ZYA-*arg*F) U169 Φ80*lac*Z ΔM15 λ^-) following the manufacturer's instructions (Bioline). Restriction fragment length polymorphism (RFLP) profiling of clones with the expected insert size was done using *Hha*I (0.2 U μL^{-1} for 3 hours at 37°C) (Promega), and profiles were visualized by electrophoresis on a 2% agarose gel. Representative inserts of unique RFLP profiles were then sequenced (Macrogen Inc., Seoul, Korea). The resulting sequences were used to evaluate the coverage and specificity of the candidate primer pairs using BLAST search (Altschul et al. 1990). Finally, the primer set *acpA*-F118 (5'- GTS GTG ATC TAC GCS GAR AA-3') / *acpA*-R589 (5'- GAY CAG GTA YTG GTG GTT-3') was selected for 454-sequencing on the soils, named after the position in the *acpA* gene in *Burkholderia cepacia* ATCC 25416.

A.3.2.4 454-sequencing of *acpA*

Prior to sequencing, DNA duplicates extracts were tested for presence of *acpA* in the samples using PCR. PCR reactions were performed in a 25 µL volume containing 1X GoTaq Reaction Buffer (including $MgCl_2$ and dNTPs), 0.5 mM of each primer and 0.6 Units of GoTaq polymerase (Promega, WI) with 1-2 ng DNA as template in an Labcycler (Sensoquest, Göttingen, Germany). The amplification reaction included an initial denaturation step of 5 min at 95°C followed by 35 cycles of a denaturation step of 30 s at 95°C, an annealing step of 30 s at 54°C and an extension step of 30 s at 72°C. A final extension step was performed for 5 min at 72°C. Quality of the amplification was assessed visually by electrophoresis on a 1.5% (w/v) agarose gel stained with ethidium bromide.

Duplicate DNA extract were then pooled and adjusted to a concentration of 20 ng/µL. Pooled extracts were sent for 454-sequencing on GS-FLX+ platform (Roche 454 Life Sciences, Branford, CT) by Research and Testing Laboratory (Lubbock, TX, http://www.research- andtesting.com). Barcoded primer design, sequencing, and initial quality filtering was performed by Research and Testing Laboratory using standard protocols (Lubbock, TX). Briefly, sequences with a quality score below 25 were trimmed. Chimeras were removed using USEARCH with a clustering at a 4% divergence (Edgar et al. 2011). Denoising was performed using Research and Testing Denoiser algorithm. It used the non-chimeric sequences and the quality scores to create consensus cluster from aligned sequences. Among each cluster, the probability of prevalence of each nucleotide was calculated and accordingly, a new quality score generated, which was then used to remove noise from the dataset.

A.3.2.5 Sequence analysis

Sequencing datasets were analyzed using MOTHUR (Schloss et al. 2009). Sequences were analyzed as nucleic acid sequences to keep the maximum information and allow accurate identification. After demultiplexing, reads including ambiguities, mismatches

with the specific primers and the barcode were removed. Reads with an average quality score below 20 were then filtered. Resulting reads were trimmed at 200 bp and 600 bp as minimum and maximum length.

Sequences were aligned using the Needleman-Wunsch global alignment algorithm as implemented in MOTHUR, using 6-mers searching and the aligned reference sequences as template. The pairwise distance matrix was calculated from the alignment. Similarity cutoff defining *acpA* OTUs was determined based on the similarity matrices at pairwise alignments of the *acpA* gene and their corresponding 16S rRNA gene of 10 strains plotted against each other (Figure A.1). The 20 strains were selected to cover best the large diversity of *acpA*. At 97% similarity of the 16S RNA gene, OTU for the *acpA* gene correspond at 72% similarity. OTU matrices were normalized to the smallest library size using the normalized.shared command in MOTHUR to allow comparison between samples, by dividing the relative abundance of each OTU by the total number of reads per sample. The normalized values were then rounded to the nearest integer.

Taxonomy assignment was performed using blastn in BLAST+ (Camacho et al. 2009) with a minimum e-value of 1e-8 to retrieve NCBI sequence identifiers (GI accession number). Subsequently, in-house Perl scripts were used to populate and query a mysql database containing the NCBI GI number and taxonomic lineage information (script written by Stefan Zoller, Genetic Diversity Centre, ETH Zurich, available on request) (Table A.2).

A.3.3 Results and Discussion

A.3.3.1 Taxonomic composition of *acpA*-harboring community in 30 soils using newly designed primers

The amplified *acpA* genes were affiliated to 25 bacterial orders among 8 different phyla (Figure A.2 and Table A.2). Between 0.1% and 1.7% of the sequences in the samples could not be classified. The *acpA* gene could be amplified from 25 out of the 30 samples.

Our results showed that the composition of the *acpA*-harboring communities varied between samples. In most soils, *acpA*-harboring *Burkholderiales* were dominant with a relative abundance ranging between 19% and 81%. *Acidobacteriales*, *Gemmatimonadetes* and *Pseudomonadales* were also key *acpA*-harboring orders in certain soils.

Our results show that the *acpA* gene is mainly spread in *Acidobacteria* (*Acidobacteriales*) and *Proteobacteria* (*Burkholderiales* and *Pseudomonadales*), which supports the findings of Costas et al. (2010) who found 88 *acpA* homologues by similarity search mainly affiliated to *Acidobacteria* and *Proteobacteria*. Moreover, the *acpA* gene was found in an additional two phyla, namely *Chlorobi* and *Gemmatimonadetes*, compared to the *acpA*-harboring genomes currently present in the IMG/M database. The diversity of the *acpA* genes amplified from the 6 grassland soils was rather large given the small number of sequences used for the primer design, which were mainly affiliated to

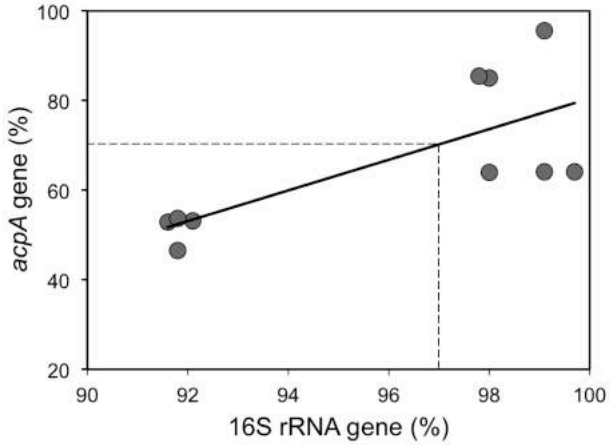


Figure A.1. Similarity percentages of pairwise alignment of the 16S rRNA gene and *acpA* gene of 8 strains. Cutoff at 97% similarity of the 16S rRNA gene corresponds to a cutoff at 72% of the *acpA* gene.

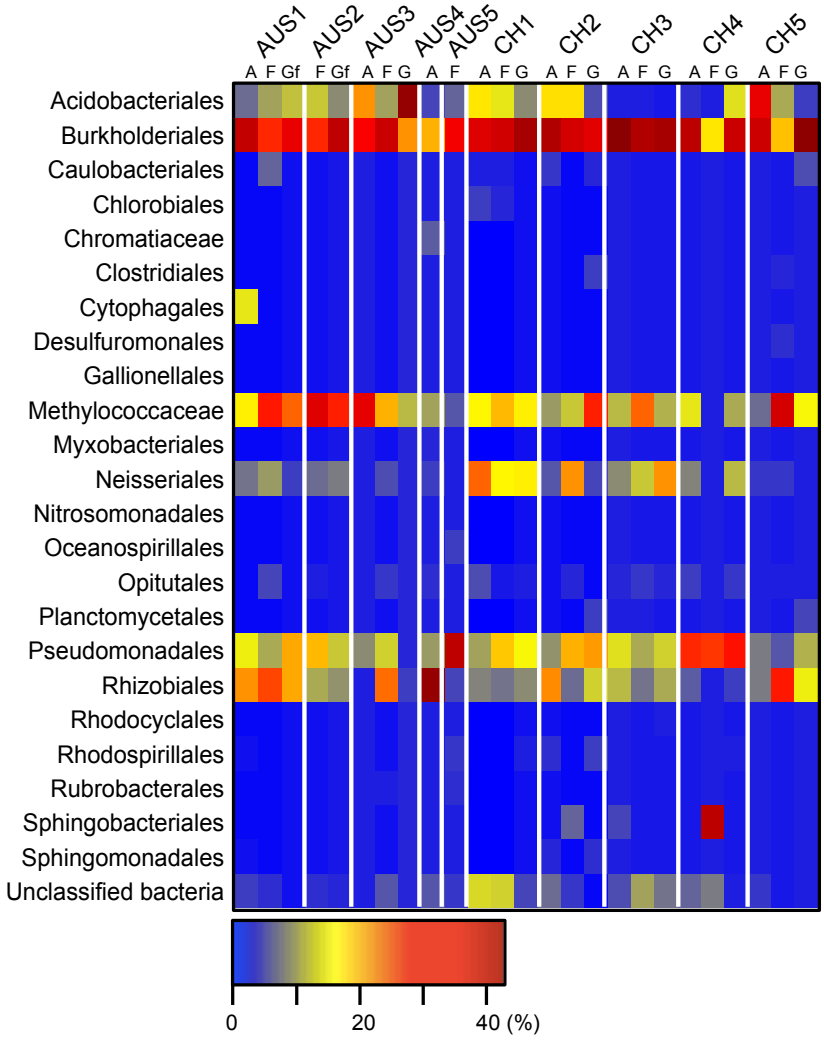


Figure A.2. Heatmap of the relative abundance of *acpA*-harboring orders.

Proteobacteria. The structure of AcpA resembles that of the PhoD and PhoX alkaline phosphatases (Felts et al. 2006).

A.3.4 Conclusion

The *acpA* gene could be amplified on 25 out of the 30 soil samples, showing that the *acpA* gene is common in soil. Additionally, the *acpA* gene was mostly found in *Actinobacteria* and *Proteobacteria*.

Table A.1. Accession numbers in EBI and NCBI databases and taxonomy of reference sequences used for the primer design of the PHOD primers.

#	Accesion number	Phylum	Order	Genus/Species
1	ENA|KCB21403	*Proteobacteria*	*Burkholderiales*	*Bordetella hinzii* L60
2	ENA|EPZ86399	*Proteobacteria*	*Burkholderiales*	*Burkholderia cenocepacia* K562
3	ENA|AEA58873	*Proteobacteria*	*Burkholderiales*	*Burkholderia gladioli* BSR3
4	ENA|ACR27360	*Proteobacteria*	*Burkholderiales*	*Burkholderia glumae* BGR1
5	ENA|EED97683	*Proteobacteria*	*Burkholderiales*	*Burkholderia multivorans* CGD1
6	ENA|EEE03148	*Proteobacteria*	*Burkholderiales*	*Burkholderia multivorans* CGD1
7	ENA|EEE09403	*Proteobacteria*	*Burkholderiales*	*Burkholderia multivorans* CGD2
8	ENA|EDO84336	*Proteobacteria*	*Burkholderiales*	*Burkholderia pseudomallei* 406e
9	ENA|EEH30718	*Proteobacteria*	*Burkholderiales*	*Burkholderia pseudomallei* Pakistan
10	ENA|EGD02585	*Proteobacteria*	*Burkholderiales*	*Burkholderia* sp. TJI49
11	ENA|AAF66062	*Proteobacteria*	*Burkholderiales*	*Burkholderia thailandensis* acid
12	ENA|AHI65670	*Proteobacteria*	*Burkholderiales*	*Burkholderia thailandensis* H0587
13	ENA|AGK49094	*Proteobacteria*	*Burkholderiales*	*Burkholderia thailandensis* MSMB121
14	ENA|AGK49321	*Proteobacteria*	*Burkholderiales*	*Burkholderia thailandensis* MSMB121
15	ENA|AAQ61187	*Proteobacteria*	*Neisseriales*	*Chromobacterium violaceum* ATCC
16	ENA|AEK62102	*Proteobacteria*	*Burkholderiales*	*Collimonas fungivorans* Ter331
17	ENA|EEI14121	*Actinobacteria*	*Cyrinebacteriales*	*Corynebacterium accolens* ATCC
18	ENA|EFM43095	*Actinobacteria*	*Cyrinebacteriales*	*Corynebacterium accolens* ATCC
19	ENA|CAI36159	*Actinobacteria*	*Cyrinebacteriales*	*Corynebacterium jeikeium* K411
20	ENA|EFQ79562	*Actinobacteria*	*Cyrinebacteriales*	*Corynebacterium pseudogenitalium* ATCC
21	ENA|EET76779	*Actinobacteria*	*Cyrinebacteriales*	*Corynebacterium tuberculostearicum* SK141
22	ENA|ABF11671	*Proteobacteria*	*Burkholderiales*	*Cupriavidus metallidurans* CH34
23	ENA|AAB06624	*Proteobacteria*	*Thiotrichales*	*Francisella tularensis* subsp.
24	ENA|CDG82200	*Proteobacteria*	*Burkholderiales*	*Janthinobacterium agaricidamnosum* NBRC
25	ENA|CDG83673	*Proteobacteria*	*Burkholderiales*	*Janthinobacterium agaricidamnosum* NBRC
26	ENA|EZP35507	*Proteobacteria*	*Burkholderiales*	*Janthinobacterium lividum* Acid
27	ENA|ELX09523	*Proteobacteria*	*Burkholderiales*	*Janthinobacterium* sp. HH01
28	ENA|AID29393	*Proteobacteria*	*Rhizobiales*	*Mesorhizobium huakuii* 7653R
29	ENA|AEB95158	*Crenoarcheaota*	*Sulfolobales*	*Metallosphaera cuprina* Ar4
30	ENA|EIM16513	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas chlororaphis* O6
31	ENA|EJL07537	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas chlororaphis* subsp.
32	ENA|EJK99898	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas chlororaphis* subsp.
33	ENA|AFJ56674	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas fluorescens* A506
34	ENA|EIK59256	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas fluorescens* SS101
35	ENA|AAY94450	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas protegens* Pf5
36	ENA|EIK67235	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas synxantha* BG33R
37	ENA|ELS43673	*Proteobacteria*	*Pseudomonadales*	*Pseudomonas syringae* pv.
38	ENA|CAD18325	*Proteobacteria*	*Burkholderiales*	*Ralstonia solanacearum* GMI1000
39	ENA|AEG71481	*Proteobacteria*	*Burkholderiales*	*Ralstonia solanacearum* Po82
40	ENA|AEG70621	*Proteobacteria*	*Burkholderiales*	*Ralstonia solanacearum* Po82
41	ENA|EFP68144	*Proteobacteria*	*Burkholderiales*	*Ralstonia* sp. 5747FAA
42	ENA|EWG06834	*Crenoarcheaota*	*Sulfolobales*	*Sulfolobales archaeon* AZ1
43	ENA|AAK42902	*Crenoarcheaota*	*Sulfolobales*	*Sulfolobus solfataricus* P2
44	ENA|KFD18434	*Proteobacteria*	*Enterobacteriales*	*Tatumella ptyseos* ATCC
45	ENA|AEA12332	*Proteobacteria*	*Thermoproteales*	*Thermoproteus uzoniensis* 76820

Table A.2. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the *acpA*-harboring community in the 30 soil samples.

Taxonomy	AUS1-A	AUS1-F	AUS1-Gf	AUS2-F	AUS2-Gf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS5-F	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-C	CH5-F	CH5-G
1. Acidobacteria	251	254	42	503	264	18	304	3542	145	20	1574	1188	315	283	729	24	0	2	0	13	0	6051	3163	257	10
1.1. Acidobacteria	251	254	42	503	264	18	304	3542	145	20	1574	1188	315	283	729	24	0	2	0	13	0	6051	3163	257	10
1.1.1. Acidobacteriales	251	254	42	503	264	18	300	3542	145	19	1574	1188	315	283	729	24	0	2	0	13	0	6051	3163	257	10
1.1.1.1. Acidobacteriaceae	240	254	42	503	264	1	296	3542	143	1	1574	1188	315	282	729	24	0	0	0	0	0	5922	3162	257	10
1.1.1.1.1. Acidobacterium	240	254	42	503	264	1	296	3542	143	1	1574	1108	315	282	729	24	0	0	0	0	0	5922	3162	257	10
1.1.1.1.2. Granulicella	0	0	0	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	0
1.1.1.2. Actinoplanaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0
1.1.1.2.1. Streptomyces	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0
1.1.1.3. Actinosynnemataceae	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1.1.1.3.1. Actinobispora	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
1.1.1.4. Microsphaeraceae	8	0	0	0	0	17	4	0	2	18	0	0	0	0	0	0	0	2	0	13	0	73	0	0	0
1.1.1.4.1. Humicoccus	8	0	0	0	0	17	4	0	2	18	0	0	0	0	0	0	0	2	0	13	0	73	0	0	0
1.1.1.5. Nocardioidaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1.1.1.5.1. Nocardioides	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1.1.2. Rubrobacterales	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.1.2.1. Rubrobacteraceae	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1.1.2.1.1. Rubrobacter	0	0	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2. Firmicutes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	2	0
2.1. Clostridia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	2	0
2.1.1. Clostridiales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	2	0
2.1.1.1. Peptococcaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2.1.1.1.1. Pelotomaculum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2.1.1.2. Cytophaga-Flexibacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0
2.1.1.2.1. Spirosoma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0
3. Bacteroidetes	1384	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1. Cytophagia	1384	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1.1. Cytophagales	1384	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.2. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the *acpA*-harboring community in the 30 soil samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Gf	AUS2-F	AUS2-Gf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS5-F	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-C	CH5-F	CH5-G
3.1.1.1. Flavobacteriaceae	1384	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1.1.1.1. Aequorivita	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.1.1.1.2. cytophcterium	1383	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3.2. Sphingobacteriia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	17	0	0	0	51	0	0	0	0
3.2.1. Sphingobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	17	0	0	0	51	0	0	0	0
3.2.1.1. Sphingobacteriaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	17	0	0	0	51	0	0	0	0
3.2.1.1.1. Pedobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0	16	0	0	0	47	0	0	0	0
3.2.1.1.2. Sphingobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	4	0	0	0	0
4. Chlorobi	0	0	0	0	0	0	0	0	0	0	97	24	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1. Chlorobia	0	0	0	0	0	0	0	0	0	0	97	24	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1. Chlorobiales	0	0	0	0	0	0	0	0	0	0	97	24	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1.1. Chlorobiacea	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1.1.1. Chlorobium/Pelodictyon	0	0	0	0	0	0	0	0	0	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1.1.2. Prosthecochloris	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1.2. Croococcaceae	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0
4.1.1.2.1. Acaryochloris	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0
5. Gemmatimonadetes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0
5.1. Gemmatimonadetes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0
6. Planctomycetes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	16
6.1. Planctomycetacia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	16
6.1.1. Planctomycetales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	16
6.1.1.1. Planctomycetaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	16
6.1.1.1.1. Singulisphaera	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	1	0	0	0	0	0	0	16
7. Proteobacteria	17805	7624	854	9243	12889	85	9458	888	53224	2551	12780	14530	13886	2315	5861	3731	7224	8610	6429	17446	3196	200	41107	4217	7226
7.1. Alphaproteobacteria	3348	2167	139	320	292	0	2105	2	42197	7	414	289	326	502	96	230	278	118	221	139	0	289	89	2504	518
7.1.1. Caulobacteriales	0	94	0	0	0	0	0	0	0	0	25	14	0	7	0	3	0	0	0	0	0	0	0	0	24
7.1.1.1. Caulobacteraceae	0	94	0	0	0	0	0	0	0	0	25	14	0	7	0	3	0	0	0	0	0	0	0	0	24
7.1.1.1.1. Caulobacter	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.2. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the *acpA*-harboring community in the 30 soil samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Gf	AUS2-F	AUS2-Gf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS5-F	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-C	CH5-F	CH5-G
7.1.1.1.2. Polymorphum	0	94	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.1.1.3. Rhizobiales	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	3	0	0	0	0	0	0	0	0	24
7.1.2. Rhizobiales	3346	2073	139	320	292	0	2105	2	42197	5	388	275	324	489	96	207	278	118	221	139	0	283	89	2504	494
7.1.2.1. Beijerinckiaceae	0	3	0	3	0	0	3	0	0	0	0	0	0	0	2	1	0	0	0	11	0	8	0	0	0
7.1.2.1.1. Beijerinckia	0	3	0	3	0	0	3	0	0	0	0	0	0	0	2	1	0	0	0	11	0	8	0	0	0
7.1.2.2. Beijerinckiaceae	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.2.2.1. Agromonas	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.2.3. Bradyrhizobiaceae	3321	1979	131	312	194	0	1935	2	42059	5	382	50	115	482	92	205	262	21	53	116	0	169	55	2504	494
7.1.2.3.1. Rhodobacillus	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.2.3.2. Mesorhizobium	3321	1979	131	312	194	0	1932	2	42059	5	382	50	115	482	92	205	262	21	53	116	0	169	55	2504	494
7.1.2.4. Phyllobacteriaceae	25	91	8	5	98	0	164	0	138	0	6	225	209	7	2	1	16	97	168	12	0	106	34	0	0
7.1.2.4.1. Rhizobium	7	91	8	5	98	0	164	0	131	0	6	225	209	7	2	1	16	97	168	12	0	106	34	0	0
7.1.2.4.2. Sinorhizobium	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.2.4.3. Starkeya	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.3. Rhodospirillales	1	0	0	0	0	0	0	0	0	2	1	0	2	4	0	14	0	0	0	0	0	6	0	0	0
7.1.3.1. Acetobacteraceae	0	0	0	0	0	0	0	0	0	2	1	0	2	2	0	8	0	0	0	0	0	6	0	0	0
7.1.3.1.1. Gluconacetobacter	0	0	0	0	0	0	0	0	0	2	1	0	2	0	0	2	0	0	0	0	0	6	0	0	0
7.1.4. Sphingomonadales	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	0	0	0	0	0	0	0	0	0
7.1.4.1. Erythrobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	0	0	0	0	0	0	0	0	0
7.1.4.1.1. Erythrobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	6	0	0	0	0	0	0	0	0	0
7.1.4.2. Sphingomonadaceae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.1.4.2.1. Rhizomonas	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2. Betaproteobacteria	11207	2626	381	3044	7920	38	5371	744	7896	1012	10386	9808	11199	1672	4471	1690	6258	6173	5642	10828	8	58961	3857	10265	936
7.2.1. Burkholderiales	10913	2380	378	2915	7710	38	5319	744	7796	1012	6927	8274	9867	1653	3291	1671	6135	5747	4538	10523	8	54994	3853	10135	936
7.2.1.1. Alcaligenaceae	0	0	0	0	0	0	0	0	0	0	17	27	1	0	0	0	1	1	27	1	0	285	2	0	0
7.2.1.1.1. Achromobacter	0	0	0	0	0	0	0	0	0	0	17	27	1	0	0	0	1	1	27	1	0	285	2	0	0
7.2.1.2. Burkholderiaceae	10825	2136	366	2728	2974	38	5004	726	7684	1012	5605	5861	5266	551	3231	1554	5804	4429	3563	6330	5	49257	988	10135	548
7.2.1.2.1. Burkholderia	8105	1531	312	2145	1244	21	4032	10	5068	996	3757	2912	1965	164	3010	721	5615	2743	1768	4704	5	39314	976	101	275

Table A.2. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the *acpA*-harboring community in the 30 soil samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Gf	AUS2-F	AUS2-Gf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS5-F	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-C	CH5-F	CH5-G
7.2.1.2.2. Cupravidus	1442	388	7	405	887	8	496	716	1542	16	1605	339	919	268	144	589	40	147	485	708	0	6764	4	289	1303
7.2.1.2.3. Pandoraea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
7.2.1.2.4. Ralstonia	1278	217	47	178	843	9	476	0	1074	0	243	2610	2382	119	77	244	149	1539	1310	918	0	3178	8	623	3970
7.2.1.3. Comamonadaceae	0	0	0	0	0	0	0	0	21	0	0	0	0	71	0	31	211	487	273	374	3	739	2846	0	384
7.2.1.3.1. Acidivorax	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	211	487	273	374	3	739	2846	0	0
7.2.1.3.2. Alicycliphilus	0	0	0	0	0	0	0	0	0	0	0	0	0	71	0	31	0	0	0	0	0	0	0	0	384
7.2.1.4. Oxalobacteraceae	88	244	12	187	4736	0	315	18	91	0	1305	2386	4600	1031	60	86	119	830	675	3818	0	4713	17	0	4
7.2.1.4.1. Collimonas	87	242	12	187	4736	0	252	18	87	0	1250	2364	4593	1030	60	83	91	827	674	3760	0	4276	0	0	4
7.2.1.4.2. Herbaspirillum	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	3	7	0	0	34	0	328	17	0	0
7.2.1.4.3. Janthinobacterium	1	2	0	0	0	0	63	0	2	0	55	22	7	1	0	0	21	3	1	24	0	109	0	0	0
7.2.2. Gallionellales	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.2.1. Gallionellaceae	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.2.1.1. Gallionella	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.3. Neisseriales	294	246	3	129	210	0	52	0	99	0	3459	1534	1332	19	1180	19	123	426	1103	305	0	3967	4	13	0
7.2.3.1. ChromobacteriaceaeA	294	246	3	129	210	0	52	0	99	0	3459	1534	1332	19	1180	19	123	426	1103	305	0	3967	4	13	0
7.2.3.1.1. Chromobacterium	294	246	3	129	210	0	52	0	99	0	3459	1534	1332	19	1180	19	123	426	1103	305	0	3967	4	13	0
7.2.4. Nitrosomonadales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.4.1. Nitrosomonadaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.4.1.1. Nitrosolobus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.2.5. Rhodocyclales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7.2.5.1. Rhodocyclaceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7.2.5.1.1. Azoarcus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7.3. Deltabacteria	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
7.3.1. Desulfuromonales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
7.3.1.1. Geobacteraceae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
7.3.1.1.1. Geobacter	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
7.3.2. Myxobacteriales	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.3.2.1. Archangiaceae	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.2. Taxonomy summary including number of reads at the phylum, class, order, family and genus levels of the *acpA*-harboring community in the 30 soil samples. (Continued)

Taxonomy	AUS1-A	AUS1-F	AUS1-Gf	AUS2-F	AUS2-Gf	AUS3-A	AUS3-F	AUS3-G	AUS4-A	AUS5-F	CH1-A	CH1-F	CH1-G	CH2-A	CH2-F	CH2-G	CH3-A	CH3-F	CH3-G	CH4-A	CH4-F	CH4-G	CH5-C	CH5-F	CH5-G
7.3.2.1.1. Stigmatella	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4. Gammaproteobacteria	3250	2831	334	5879	4677	47	1982	142	3127	1532	1980	4433	2361	141	1294	1811	688	2319	566	6479	23	36950	164	3884	772
7.4.1. Chromatiaceae	0	0	0	0	0	0	0	0	316	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.1.1. Chromatiaceae	0	0	0	0	0	0	0	0	316	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.1.1.1. Chromatium	0	0	0	0	0	0	0	0	316	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.2. Methylococcaceae	1808	2543	197	4523	4059	45	1429	142	1551	10	1443	2290	1287	77	335	1182	282	2013	223	1126	0	3224	64	3843	532
7.4.2.1. Methylococcaceae	1808	2543	197	4523	4059	45	1429	142	1551	10	1443	2290	1287	77	335	1182	282	2013	223	1126	0	3224	64	3843	532
7.4.2.1.1. Methylococcus	1808	2543	197	4523	4059	45	1429	142	1551	10	1443	2290	1287	77	335	1182	282	2013	223	1126	0	3224	64	3843	532
7.4.3. Oceanospirillales	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.3.1. Alcanivoracaceae	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.3.1.1. Alcanivorax	0	0	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7.4.4. Pseudomonadales	1442	288	137	1356	618	2	553	0	1260	1518	536	2143	1074	64	959	629	406	306	343	5353	23	33726	100	41	240
7.4.4.1. Pseudomonadaceae	1442	288	137	1346	618	2	549	0	1255	1512	524	2127	305	61	952	626	406	280	232	5348	23	31611	97	21	238
7.4.4.1.1. Cellvibrio	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	8	0	0	0	0	0	0	0	0	0
7.4.4.1.2. Pseudomonas	1442	288	137	1346	618	2	549	0	1255	1512	524	2127	305	38	952	618	406	280	232	5348	23	31611	97	21	238
7.4.4.2. Lysobacteraceae	0	0	0	10	0	0	4	0	5	6	12	16	769	3	7	3	0	26	111	5	0	2115	3	20	2
7.4.4.2.1. Agrobacterium	0	0	0	0	0	0	0	0	1	0	0	15	2	0	0	1	0	5	1	4	0	0	0	0	0
7.4.4.2.2. Dyella	0	0	0	0	0	0	4	0	0	4	4	0	228	0	0	0	0	2	9	0	0	955	0	0	0
7.4.4.2.3. Frateuria	0	0	0	0	0	0	0	0	0	2	1	0	2	2	0	2	0	0	0	0	0	17	0	2	2
7.4.4.2.4. Rhodanobacter	0	0	0	10	0	0	0	0	4	0	7	1	537	1	7	0	0	19	101	1	0	1143	3	18	0
8. Verrucomicrobia	0	41	0	2	0	0	28	0	21	0	122	9	5	0	8	0	1	14	3	37	0	128	0	1	0
8.1. Opitutae	0	41	0	2	0	0	28	0	21	0	122	9	5	0	8	0	1	14	3	37	0	128	0	1	0
8.1.1. Opitutales	0	41	0	2	0	0	28	0	21	0	122	9	5	0	8	0	1	14	3	37	0	128	0	1	0
8.1.1.1. Opitutaceae	0	41	0	2	0	0	28	0	21	0	122	9	5	0	8	0	1	14	3	37	0	128	0	1	0
8.1.1.1.1. unclassified Opitutaceae	0	41	0	2	0	0	28	0	21	0	122	9	5	0	8	0	1	14	3	37	0	128	0	1	0
9. Unclassidied bacteria	82	17	0	11	5	0	67	0	255	2	995	929	54	34	24	0	21	277	85	152	1	11	7	0	0

BIBLIOGRAPHY

Acinas, S., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. Polz (2005). PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and Environmental Microbiology* 71, pp. 8966–8969.

Adams, M., M. Gómez-García, A. Grossman, and D. Bhaya (2008). Phosphorus deprivation responses and phosphonate utilization in a thermophilic *Synechococcus* sp. from microbial mats. *Journal of Bacteriology* 190, pp. 8171–8184.

Alef, K. and P. Nannipieri (1995). *Methods in Applied Soil Microbiology and Biochemistry*. Academic Press, London.

Allison, S., M. Weintraub, T. Gartner, and M. Waldrop (2011). Evolutionary-economic principles as regulators of soil enzyme production and ecosystem function. In: *Soil Enzymology*. Springer, pp. 229–243.

Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, pp. 403–410.

Anderson, J. and J. Ingram (1993). *Tropical soil biology and fertility: a handbook of methods*. CAB international Wallingford, Oxfordshire.

Ansai, T. et al. (1998). Purification and characterization of alkaline phosphatase containing phosphotyrosyl phosphatase activity from the bacterium *Prevotella intermedia*. *FEBS Letters* 428, pp. 157–160.

Apel, A. K., A. Sola-Landa, A. Rodriguez-Garcia, and J. F. Martin (2007). Phosphate control of *phoA, phoC* and *phoD* gene expression in *Streptomyces coelicolor* reveals significant differences in binding of PhoP to their promoter regions. *Microbiology* 153, pp. 3527–3537.

Ashby, M. (2006). Distribution, structure and diversity of "bacterial" genes encoding two-component proteins in the *Euryarchaeota*. *Archaea* 2, pp. 11–30.

Bachar, A. et al. (2010). Soil microbial abundance and diversity along a low precipitation gradient. *Microbial Ecology* 60, pp. 453–461.

Baker, G., J. Smith, and D. Cowan (2003). Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods* 55, pp. 541–555.

Baldrian, P. et al. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME Journal* 6, pp. 248–258.

Balota, E., I. Yada, H. Amaral, A. Nakatani, R. Dick, and M. Coyne (2014). Long-term land-use influences soil microbial biomass P and S, phosphatase and arylsulfatase activities, and S mineralization in a Brasilian Oxisol. *Land Degradation and Development* 25, pp. 397–406.

Baltar, F., J. Arístegui, J. Gasol, T. Yokokawa, and G. Herndl (2013). Bacterial versus archaeal origin of extracellular enzymatic activity in the Northeast Atlantic deep waters. *Microbial Ecology* 65, pp. 277–288.

Beardall, J., E. Young, and S. Roberts (2001). Approaches for determining phytoplankton nutrient limitation. *Aquatic Sciences* 63, pp. 44–69.

Beauregard, M., C. Hamel, and M. St-Arnaud (2010). Long-term phosphorus fertilization impacts soil fungal and bacterial diversity but not AM fungal community in alfalfa. *Microbial Ecology* 59, pp. 379–389.

Beazley, M., R. Martinez, S. Webb, P. Sobecky, and M. Taillefert (2011). The effect of pH and natural microbial phosphatase activity on the speciation of uranium in subsurface soils. *Geochimica et Cosmochimica Acta* 75, pp. 5648–5663.

Berg, G. and K. Smalla (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology* 68, pp. 1–13.

Bergkemper, F. et al. (2015). Phosphorus depletion in forest soils shapes bacterial communities towards phosphorus recycling systems. *Environmental Microbiology*.

Birkhofer, K. et al. (2012). General relationships between abiotic soil properties and soil biota across spatial scales and different land-use types. *PloS ONE*, e43292.

Bissett, A., A. Richardson, G. Baker, and P. Thrall (2011). Long-term land use effects on soil microbial community structure and function. *Applied Soil Ecology* 51, pp. 66–78.

Blades, A., Y. Ho, and P. Kebarle (1996). Hydration in the gas phase of the orthophosphate anion,$(HO)_2PO_2^-$, and the conversion of the orthophosphate to the metaphosphate, PO_3^-, Ion. *Journal of the American Chemical Society* 118, pp. 196–201.

Blagodatskaya, E. and T.-H. Anderson (1998). Interactive effects of pH and substrate quality on the fungal-to-bacterial ratio and qCO_2 of microbial communities in forest soils. *Soil Biology and Biochemistry* 30, pp. 1269–1274.

Blake, R. E., J. R. O'Neil, and A. V. Surkov (2005). Biogeochemical cycling of phosphorus: insights from oxygen isotope effects of phosphoenzymes. *American Journal of Science* 305, pp. 596–620.

Boulanger, R. and E. Kantrowitz (2003). Characterization of a monomeric *Escherichia coli* alkaline phosphatase formed upon a single amino acid substitution. *Journal of Biological Chemistry* 278, pp. 23497–23501.

Bowman, R. and J. Moir (1993). Basic EDTA as an extractant for soil organic phosphorus. *Soil Science Society of America Journal* 57, pp. 1516–1518.

Brown, J. and W. Doolittle (1997). Archaea and the prokaryote-to-eukaryote transition. *Microbiology and Molecular Biology Reviews* 61, pp. 456–502.

Buckley, D. H., V. Huangyutitham, S.-F. Hsu, and T. A. Nelson (2007). Stable isotope probing with $^{15}N_2$ reveals novel noncultivated diazotrophs in soil. *Applied and environmental microbiology* 73, pp. 3196–3204.

Bünemann, E. (2015). Assessment of gross and net mineralization rates of soil organic phosphorus - A review. *Soil Biology and Biochemistry* 89, pp. 82–98.

Bünemann, E., P. Marschner, A. McNeill, and M. J. McLaughlin (2007). Measuring rates of gross and net mineralisation of organic phosphorus in soils. *Soil Biology and Biochemistry* 39, pp. 900–913.

Bünemann, E. et al. (2012). Rapid microbial phosphorus immobilization dominates gross phosphorus fluxes in a grassland soil with low inorganic phosphorus availability. *Soil Biology and Biochemistry* 51, pp. 84–95.

Burke, I., C. Yonker, W. Parton, C. Cole, D. Schimel, and K. Flach (1989). Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. *Soil Science Society of America Journal* 53, pp. 800–805.

Burns, R. (1978). Enzyme activity in soil: some theoretical and practical considerations. In: *Soil enzymes*. Academic Press London, pp. 295–340.

Burns, R. (1982). Enzyme activity in soil: location and a possible role in microbial ecology. *Soil Biology and Biochemistry* 14, pp. 423–427.

Burns, R. et al. (2013). Soil enzymes in a changing environment: current knowledge and future directions. *Soil Biology and Biochemistry* 58, pp. 216–234.

Burtnick, M., A. Bolton, P. Brett, D. Watanabe, and D. Woods (2001). Identification of the acid phosphatase (*acpA*) gene homologues in pathogenic and non-pathogenic *Burkholderia* spp. facilitates TnphoA mutagenesis. *Microbiology* 147, pp. 111–120.

Camacho, C. et al. (2009). BLAST+: architecture and applications. *BMC bioinformatics* 10, p. 421.

Cerling, T. (1984). The stable isotopic composition of modern soil carbonate and its relationship to climate. *Earth and Planetary Science Letters* 71, pp. 229–240.

Chao, A. and T.-J. Shen (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environmental and Ecological Statistics* 10, pp. 429–443.

Chao, X., Y. Jia, C. Cooper, F. Shields Jr, and S. Wang (2006). Development and application of a phosphorus model for a shallow oxbow lake. *Journal of Environmental Engineering* 132, pp. 1498–1507.

Chhabra, S., D. Brazil, J. Morrissey, J. Burke, F. O'Gara, and D. Dowling (2013). Fertilization management affects the alkaline phosphatase bacterial community in barley rhizosphere soil. *Biology and Fertility of Soils* 49, pp. 31–39.

Cho, J.-C., K. Vergin, R. Morris, and S. Giovannoni (2004). *Lentisphaera araneosa* gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. *Environmental Microbiology* 6, pp. 611–621.

Condron, L., B. Turner, and B. Cade-Menun (2005). Chemistry and dynamics of soil organic phosphorus. *Agronomy* 46, pp. 87–121.

Costas, M. et al. (2010). CGDEase, a *Pseudomonas fluorescens* protein of the PLC/APase superfamily with CDP-ethanolamine and (dihexanoyl) glycerophosphoethanolamine hydrolase activity induced by osmoprotectants under phosphate-deficient conditions. *Molecular Microbiology* 78, pp. 1556–1576.

Coveley, S., M. Elshahed, and N. Youssef (2015). Response of the rare biosphere to environmental stressors in a highly diverse ecosystem (Zodletone spring, OK, USA). *PeerJ* 3, e1182.

Cruz, A., C. Hamel, K. Hanson, F. Selles, and R. Zentner (2009). Thirty-seven years of soil nitrogen and phosphorus fertility management shapes the structure and function of the soil microbial community in a Brown Chernozem. *Plant and Soil* 315, pp. 173–184.

Cui, H., Y. Zhou, Z. Gu, H. Zhu, S. Fu, and Q. Yao (2015). The combined effects of cover crops and symbiotic microbes on phosphatase gene and organic phosphorus hydrolysis in subtropical orchard soils. *Soil Biology and Biochemistry* 82, pp. 119–126.

Culman, S., R. Bukowski, H. Gauch, H. Cadillo-Quiroz, and D. Buckley (2009). T-REX: software for the processing and analysis of T-RFLP data. *BMC Bioinformatics* 10, p. 171.

Dai, J. et al. (2014). Recovery of novel alkaline phosphatase-encoding genes (*phoX*) from eutrophic Lake Taihu. *Canadian Journal of Microbiology* 60, pp. 167–171.

Danon, M., I. Franke-Whittle, H. Insam, Y. Chen, and Y. Hadar (2008). Molecular analysis of bacterial community succession during prolonged compost curing. *FEMS Microbiology Ecology* 65, pp. 133–144.

Dassa, E. and P. Boquet (1985). Identification of the gene *appA* for the acid phosphatase (pH optimum 2.5) of *Escherichia coli*. *Molecular and General Genetics* 200, pp. 68–73.

Dassa, J., C. Marck, and P. Boquet (1990). The complete nucleotide sequence of the *Escherichia coli* gene *appA* reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase. *Journal of Bacteriology* 172, pp. 5497–5500.

De Buck, E., E. Lammertyn, and J. Anné (2008). The importance of the twin-arginine translocation pathway for bacterial virulence. *Trends in Microbiology* 16, pp. 442–453.

Dennis, D. et al. (2001). Tularemia as a biological weapon: medical and public health management. *Jama* 285, pp. 2763–2773.

Dequiedt, S. et al. (2011). Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Global Ecology and Biogeography* 20, pp. 641–652.

Dick, W., L. Cheng, and P. Wang (2000). Soil acid and alkaline phosphatase activity as pH adjustment indicators. *Soil Biology and Biochemistry* 32, pp. 1915–1919.

Dickson, A. (1993). The measurement of sea water pH. *Marine Chemistry* 44, pp. 131–142.

Dieffenbach, C., T. Lowe, and G. Dveksler (1993). General concepts for PCR primer design. *PCR Methods Applications* 3, S30–S37.

Dietl, W. (1995). Wandel der Wiesenvegetation im Schweizer Mittelland. *Zeitschrift für Ökologie und Naturschutz* 4, pp. 239–249.

Dowd, W. (2012). Challenges for biological interpretation of environmental proteomics data in non-model organisms. *Integrative and Comparative Biology* 52, pp. 705–720.

Drenovsky, R., D. Vo, K. Graham, and K. Scow (2004). Soil water content and organic carbon availability are major determinants of soil microbial community composition. *Microbial Ecology* 48, pp. 424–430.

Dyhrman, S. T. and K. C. Ruttenberg (2006). Presence and regulation of alkaline phosphatase activity in eukaryotic phytoplankton from the coastal ocean: Implications for dissolved organic phosphorus remineralization. *Limnology and Oceanography* 51, pp. 1381–1390.

Eder, S., L. Shi, K. Jensen, K. Yamane, and F. Hulett (1996). A *Bacillus subtilis* secreted phosphodiesterase/alkaline phosphatase is the product of a Pho regulon gene, *phoD*. *Microbiology* 142, pp. 2041–2047.

Edgar, R. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, pp. 1792–1797.

Edgar, R., B. Haas, J. Clemente, C. Quince, and R. Knight (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, pp. 2194–2200.

Fardeau, J. (1993). Le phosphore assimilable des sols: sa représentation par un modèle fonctionnel à plusieurs compartiments. *Agronomie* 13, pp. 317–331.

Feller, C., E. Frossard, and M. Brassard (1994). Activité phosphatasique de quelques sols tropicaux à argile 1: 1. Répartition dans les fractions granulométriques. *Canadian Journal of Soil Science* 74, pp. 121–129.

Felts, R., T. Reilly, and J. Tanner (2006). Structure of *Francisella tularensis* AcpA. *Journal of Biological Chemistry* 281, pp. 30289–30298.

Fernandez, N. and B. Kidney (2007). Alkaline phosphatase: beyond the liver. *Veterinary Clinical Pathology* 36, pp. 223–233.

Fierer, N. and R. Jackson (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* 103, pp. 626–631.

Fraser, T. D., D. H. Lynch, E. Bent, M. H. Entz, and K. E. Dunfield (2015a). Soil bacterial *phoD* gene abundance and expression in response to applied phosphorus and long-term management. *Soil Biology and Biochemistry*, pp. 137–147.

Fraser, T., D. H. Lynch, M. H. Entz, and K. E. Dunfield (2015b). Linking alkaline phosphatase activity with bacterial *phoD* gene abundance in soil from a long-term management trial. *Geoderma* 257, pp. 115–122.

Frossard, E., M. Brossard, M. Hedley, and A. Metherell (1995). Reactions controlling the cycle of P in soils. In: *Phosphorus in the global environment. Transfers, cycles and management*. New York: John Wiley and Sons, pp. 107–137.

Frossard, E., E. Bünemann, J. Jansa, A. Oberson, and C. Feller (2009). Concepts and practices of nutrient management in agro-ecosystems: can we draw lessons from history to design future sustainable agricultural production systems. *Die Bodenkultur* 60, pp. 43–60.

Frossard, E., P. Demaria, S. Sinaj, and M. Schärer (2014). A flow-through reactor to assess potential phosphate release from agricultural soils. *Geoderma* 219, pp. 125–135.

Frossard, E., D. López-Hernández, and M. Brossard (1996). Can isotopic exchange kinetics give valuable information on the rate of mineralization of organic phosphorus in soils? *Soil Biology and Biochemistry* 28, pp. 857–864.

Galperin, M. and M. Jedrzejas (2001). Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes. *Proteins: Structure, Function, and Bioinformatics* 45, pp. 318–324.

Galperin, M., K. Makarova, Y. Wolf, and E. Koonin (2014). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, gku1223.

Gardes, M. and T. Bruns (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2, pp. 113–118.

Garg, S. and G. Bahl (2008). Phosphorus availability to maize as influenced by organic manures and fertilizer P associated phosphatase activity in soils. *Bioresource Technology* 99, pp. 5773–5777.

George, T., A. Richardson, P. Hadobas, and R. Simpson (2004). Characterization of transgenic *Trifolium subterraneum* L. which expresses *phyA* and releases extracellular phytase: growth and P nutrition in laboratory media and soil. *Plant, Cell and Environment* 27, pp. 1351–1361.

George, T., A. Richardson, and R. Simpson (2005). Behaviour of plant-derived extracellular phytase upon addition to soil. *Soil Biology and Biochemistry* 37, pp. 977–988.

Girvan, M., J. Bullimore, J. Pretty, A. Osborn, and A. Ball (2003). Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. *Applied and Environmental Microbiology* 69, pp. 1800–1809.

Goldman, S., K. Hecht, H. Eisenberg, and M. Mevarech (1990). Extracellular Ca^{2+}-dependent inducible alkaline phosphatase from extremely halophilic archaebacterium *Haloarcula marismortui*. *Journal of Bacteriology* 172, pp. 7065–7070.

Golovan, S., G. Wang, J. Zhang, and C. Forsberg (1999). Characterization and overproduction of the *Escherichia coli appA* encoded bifunctional enzyme that exhibits both phytase and acid phosphatase activities. *Canadian Journal of Microbiology* 46, pp. 59–71.

Gomez, P. and L. Ingram (1995). Cloning, sequencing and characterization of the alkaline phosphatase gene *phoD* from *Zymomonas mobilis*. *FEMS Microbiology Letters* 125, pp. 237–245.

Good, I. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* 40, pp. 237–264.

Gotelli, N. and R. Colwell (2011). Estimating species richness. *Biological diversity: frontiers in measurement and assessment* 12, pp. 39–54.

Griffiths, R., B. Thomson, P. James, T. Bell, M. Bailey, and A. Whiteley (2011). The bacterial biogeography of British soils. *Environmental Microbiology* 13, pp. 1642–1654.

Guggenberger, G., B. Christensen, G. Rubaek, and W. Zech (1996). Land-use and fertilization effects on P forms in two European soils: resin extraction and ^{31}P-NMR analysis. *European Journal of Soil Science* 47, pp. 605–614.

Gulvik, C., T. Effler, S. Wilhelm, and A. Buchan (2012). De-MetaST-BLAST: A tool for the validation of degenerate primer sets and data mining of publicly available metagenomes. *PloS ONE*, e50362.

Guo, L. and R. Gifford (2002). Soil carbon stocks and land use change: a meta analysis. *Global change biology* 8, pp. 345–360.

Haas, H., B. Redl, E. Friedlin, and G. Stöffler (1992). Isolation and analysis of the *Penicillium chrysogenum phoA* gene encoding a secreted phosphate-repressible acid phosphatase. *Gene* 113, pp. 129–133.

Hallmann, A. (1999). Enzymes in the extracellular matrix of Volvox: an inducible, calcium-dependent phosphatase with a modular composition. *Journal of Biological Chemistry* 274, pp. 1691–1697.

Hammer, Ø., D. Harper, and P. Ryan (2001). PAST-Palaeontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4, 9pp.

Harrison, A. (1987). *Soil organic phosphorus: a review of world literature*. CAB International Wallingford.

Hayatsu, M., K. Tago, and M. Saito (2008). Various players in the nitrogen cycle: diversity and functions of the microorganisms involved in nitrification and denitrification. *Soil Science and Plant Nutrition* 54, pp. 33–45.

He, J.-Z. et al. (2007). Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices. *Environmental Microbiology* 9, pp. 2364–2374.

Hengge, A. (2005). Mechanistic studies on enzyme-catalyzed phosphoryl transfer. *Advances in Physical Organic Chemistry* 40, pp. 49–108.

Hermans, C., J. Hammond, P. White, and N. Verbruggen (2006). How do plants respond to nutrient shortage by biomass allocation? *Trends in Plant Science* 11, pp. 610–617.

Hinsinger, P. (2001). Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: a review. *Plant and Soil* 237, pp. 173–195.

Huang, Y., B. Niu, Y. Gao, L. Fu, and W. Li (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26, pp. 680–682.

Huguenin-Elie, O., R. Gago, C. Stutz, A. Lüscher, and W. Kessler (2006). Long-term effects of fertilisation on herbage composition, yield and quality of an *Arrhenatherion*-type meadow. *Grassland Science in Europe* 11, pp. 550–552.

IUSS (2014). World reference base for soil resources. *World Soil Resources Report* 103.

Jaisi, D. P., R. E. Blake, and R. K. Kukkadapu (2010). Fractionation of oxygen isotopes in phosphate during its interactions with iron oxides. *Geochimica et Cosmochimica Acta* 74, pp. 1309–1319.

Jansson, M., H. Olsson, and K. Pettersson (1988). Phosphatases; origin, characteristics and function in lakes. In: *Phosphorus in Freshwater Ecosystems*. Springer, pp. 157–175.

Jarosch, K., A. Doolette, R. Smernik, F. Tamburini, E. Frossard, and E. Bünemann (2015). Characterisation of soil organic phosphorus in NaOH-EDTA extracts: A comparison of ^{31}P NMR spectroscopy and enzyme addition assays. *Soil Biology and Biochemistry* 91, pp. 298–309.

Jorquera, M., O. Martínez, L. Marileo, J. Acuña, S. Saggar, and M. Mora (2014). Effect of nitrogen and phosphorus fertilization on the composition of rhizobacterial communities of two Chilean Andisol pastures. *World Journal of Microbiology and Biotechnology* 30, pp. 99–107.

Kageyama, H., K. Tripathi, A. Rai, S. Cha-um, R. Waditee-Sirisattha, and T. Takabe (2011). An alkaline phosphatase/phosphodiesterase, PhoD, induced by salt stress and secreted out of the cells of *Aphanothece halophytica*, a halotolerant cyanobacterium. *Applied and Environmental Microbiology* 77, pp. 5178–5183.

Kathuria, S. and A. Martiny (2011). Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium *Prochlorococcus* and other ocean bacteria. *Environmental Microbiology* 13, pp. 74–83.

Kobori, H. and N. Taga (1979). Phosphatase activity and its role in the mineralization of organic phosphorus in coastal sea water. *Journal of Experimental Marine Biology and Ecology* 36, pp. 23–39.

Kolowith, L., E. Ingall, and R. Benner (2001). Composition and cycling of marine organic phosphorus. *Limnology and Oceanography* 46, pp. 309–320.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15, pp. 259–263.

Kouno, K., Y. Tuchiya, and T. Ando (1995). Measurement of soil microbial biomass phosphorus by an anion exchange membrane method. *Soil Biology and Biochemistry* 27, pp. 1353–1357.

Lane, D. (1991). 16S/23S rRNA sequencing. In: *Nucleic acid techniques in bacterial systematics*. Wiley, pp. 125–175.

Lauber, C., M. Hamady, R. Knight, and N. Fierer (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* 75, pp. 5111–5120.

Lauber, C., M. Strickland, M. Bradford, and N. Fierer (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biology and Biochemistry* 40, pp. 2407–2415.

Lee, Y. et al. (2005). The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Research* 33, pp. D71–D74.

Legendre, P. and E. Gallagher (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, pp. 271–280.

Lehtovirta, L., J. Prosser, and G. Nicol (2009). Soil pH regulates the abundance and diversity of Group 1.1 c *Crenarchaeota*. *FEMS Microbiology Ecology* 70, pp. 367–376.

Li, H., M. Veldhuis, and A. Post (1998). Alkaline phosphatase activities among planktonic communities in the northern Red Sea. *Marine Ecology Progress Series* 173, pp. 107–115.

Liang, Y. and R. Blake (2006). Oxygen isotope signature of Pi regeneration from organic compounds by phosphomonoesterases and photooxidation. *Geochimica et Cosmochimica Acta* 70, pp. 3957–3969.

Liang, Y. and R. E. Blake (2007). Oxygen isotope fractionation between apatite and aqueous-phase phosphate: 20–45 °C. *Chemical geology* 238, pp. 121–133.

Liao, R.-Z. and P. Siegbahn (2015). Phosphate hydrolysis by the Fe^{2+}—Ca^{3+}-dependent alkaline phosphatase PhoX: Mechanistic insights from DFT calculations. *Inorganic Chemistry* 54, pp. 11941–11947.

Liebisch, F., E. Bünemann, O. Huguenin-Elie, B. Jeangros, E. Frossard, and A. Oberson (2013). Plant phosphorus nutrition indicators evaluated in agricultural grasslands managed at different intensities. *European Journal of Agronomy* 44, pp. 67–77.

Limansky, A. and A. Viale (2002). Can composition and structural features of oligonucleotides contribute to their wide-scale applicability as random PCR primers in mapping bacterial genome diversity? *Journal of Microbiological Methods* 50, pp. 291–297.

Liu, L., P. Gundersen, T. Zhang, and J. Mo (2012). Effects of phosphorus addition on soil microbial biomass and community composition in three forest types in tropical China. *Soil Biology and Biochemistry* 44, pp. 31–38.

Löffler, U., H. Cypionka, and J. Löffler (2008). Soil microbial activity along an arctic-alpine altitudinal gradient from a seasonal perspective. *European Journal of Soil Science* 59, pp. 842–854.

Lord, N., C. Kaplan, P. Shank, C. Kitts, and S. Elrod (2002). Assessment of fungal diversity using terminal restriction fragment (TRF) pattern analysis: comparison of 18S and ITS ribosomal regions. *FEMS Microbiology Ecology* 42, pp. 327–337.

Lueders, T. and M. Friedrich (2000). Archaeal population dynamics during sequential reduction processes in rice field soil. *Applied and Environmental Microbiology* 66, pp. 2732–2742.

Luo, H., R. Benner, R. Long, and J. Hu (2009). Subcellular localization of marine bacterial alkaline phosphatases. *Proceedings of the National Academy of Sciences* 106, pp. 21219–21223.

Luo, M. et al. (2010). Characterization of a monomeric heat-labile classical alkaline phosphatase from *Anabaena* sp. PCC7120. *Biochemistry* 75, pp. 655–664.

Mäder, P., A. Fliessbach, D. Dubois, L. Gunst, P. Fried, and U. Niggli (2002). Soil fertility and biodiversity in organic farming. *Science* 296, pp. 1694–1697.

Mander, C., S. Wakelin, S. Young, and M. O'Callaghan (2012). Incidence and diversity of phosphate-solubilising bacteria are linked to phosphorus status in grassland soils. *Soil Biology and Biochemistry* 44, pp. 93–101.

Marchler-Bauer, A. et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39, pp. D225–D229.

Markowitz, V. et al. (2012). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research* 40, pp. D123–D129.

Marx, M.-C., M. Wood, and S. Jarvis (2001). A microplate fluorimetric assay for the study of enzyme diversity in soils. *Soil Biology and Biochemistry* 33, pp. 1633–1640.

Matos, C. F. et al. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnology progress* 30, pp. 281–290.

McKane, R. et al. (2002). Resource-based niches provide a basis for plant species diversity and dominance in arctic tundra. *Nature* 415, pp. 68–71.

Menzel, P., P. Stadler, and J. Gorodkin (2011). maxAlike: maximum likelihood-based sequence reconstruction with application to improved primer design for unknown sequences. *Bioinformatics* 27, pp. 317–325.

Mészáros, É., R. Sipos, R. Pál, C. Romsics, and K. Márialigeti (2013). Stimulation of trichloroethene biodegradation in anaerobic three-phase microcosms. *International Biodeterioration and Biodegradation* 84, pp. 126–133.

Millan, J. (1986). Molecular cloning and sequence analysis of human placental alkaline phosphatase. *Journal of Biological Chemistry* 261, pp. 3112–3115.

Monson, R. et al. (2006). Winter forest soil respiration controlled by climate and microbial community composition. *Nature* 439, pp. 711–714.

Moseley, J., C.-W. Chang, and A. Grossman (2006). Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in *Chlamydomonas reinhardtii*. *Eukaryotic Cell* 5, pp. 26–44.

Moura, R., J. Martín, A. Martín, and P. Liras (2001). Substrate analysis and molecular cloning of the extracellular alkaline phosphatase of *Streptomyces griseus*. *Microbiology* 147, pp. 1525–1533.

Van Mourik, A., N. Bleumink-Pluym, L. van Dijk, J. van Putten, and M. Wösten (2008). Functional analysis of a *Campylobacter jejuni* alkaline phosphatase secreted via the Tat export machinery. *Microbiology* 154, pp. 584–592.

Mueller, R. et al. (2015). Metagenome sequencing of a coastal marine microbial community from Monterey Bay, California. *Genome Announcements* 3, e00341–15.

Mullaney, E. and A. Ullah (2003). The term phytase comprises several different classes of enzymes. *Biochemical and biophysical research communications* 312, pp. 179–184.

Nacke, H. et al. (2011). Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PloS ONE* 6, e17000.

Nadeau, J., R. Qualls, R. Nowak, and R. Blank (2007). The potential bioavailability of organic C, N, and P through enzyme hydrolysis in soils of the Mojave Desert. *Biogeochemistry* 82, pp. 305–320.

Nannipieri, P., L. Giagnoni, L. Landi, and G. Renella (2011). Role of phosphatase enzymes in soil. In: *Phosphorus in action*. Springer, pp. 215–243.

Narisawa, S., L. Huang, A. Iwasaki, H. Hasegawa, D. Alpers, and J. Millán (2003). Accelerated fat absorption in intestinal alkaline phosphatase knockout mice. *Molecular and Cellular Biology* 23, pp. 7525–7530.

Nelson, K. et al. (1999). Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, pp. 323–329.

Nicholls, H. and D. Osborn (1979). Bacterial stress: prerequisite for biological removal of phosphorus. *Water Pollution Control Federation*, pp. 557–569.

Nicol, G., S. Leininger, C. Schleper, and J. Prosser (2008). The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environmental Microbiology* 10, pp. 2966–2978.

Oberson, A. and E. Joner (2005). Microbial turnover of phosphorus in soil. In: *Organic phosphorus in the environment*. CABI, Wallingford, pp. 133–164.

Oberson, A. et al. (2006). Improving phosphorus fertility in tropical soils through biological interventions.

O'Brien, H., J. Parrent, J. Jackson, J.-M. Moncalvo, and R. Vilgalys (2005). Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71, pp. 5544–5550.

O'Donnell, A., S. Colvan, E. Malosso, and S. Supaphol (2005). Twenty years of molecular analysis of bacterial communities in soils and what have we learned about function. In: *Biological diversity and function in soils*. Cambridge University Press, Cambridge, pp. 44–56.

Oehl, F., E. Frossard, A. Fliessbach, D. Dubois, and A. Oberson (2004). Basal organic phosphorus mineralization in soils under different farming systems. *Soil Biology and Biochemistry* 36, pp. 667–675.

Oehl, F., A. Oberson, S. Sinaj, and E. Frossard (2001). Organic phosphorus mineralization studies using isotopic dilution techniques. *Soil Science Society of America Journal* 65, pp. 780–787.

Ohno, T. and L. Zibilske (1991). Determination of low concentrations of phosphorus in soil extracts using malachite green. *Soil Science Society of America Journal* 55, pp. 892–895.

Pál, C., B. Papp, and M. Lercher (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics* 37, pp. 1372–1375.

Paragas, V., J. Kramer, C. Fox, R. Haugland, and V. Singer (2002). The ELF®-97 phosphatase substrate provides a sensitive, photostable method for labelling cytological targets. *Journal of Microscopy* 206, pp. 106–119.

Parton, W., D. Schimel, C. Cole, and D. Ojima (1987). Analysis of factors controlling soil organic matter levels in Great Plains grasslands. *Soil Science Society of America Journal* 51, pp. 1173–1179.

Pastor, J. and W. Post (1986). Influence of climate, soil moisture, and succession on forest carbon and nitrogen cycles. *Biogeochemistry* 2, pp. 3–27.

Paytan, A. and K. McLaughlin (2007). The oceanic phosphorus cycle. *Chemical Reviews* 107, pp. 563–576.

Pettit, N., L. Gregory, R. Freedman, and R. Burns (1977). Differential stabilities of soil enzymes: assay and properties of phosphatase and arylsulphatase. *Biochimica et Biophysica Acta - Enzymology* 485, pp. 357–366.

Philipp, A., O. Huguenin-Elie, R. Flisch, R. Gago, C. Stutz, and W. Kessler (2004). Einfluss der Phosphordüngung auf eine Fromentalwiese. *Agrarforschung* 11, pp. 86–91.

Philippe, H. et al. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* 9, p. 402.

Philippot, L. et al. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology* 8, pp. 523–529.

Piddington, C. et al. (1993). The cloning and sequencing of the genes encoding phytase (*phy*) and pH 2.5-optimum acid phosphatase (*aph*) from *Aspergillus niger* var. *awamori*. *Gene* 133, pp. 55–62.

Plante, A. (2007). Soil biogeochemical cycling of inorganic nutrients and metals. In: *Soil microbiology, ecology and biochemistry*. Oxford, Elsevier.

Plassard, C. and B. Dell (2010). Phosphorus nutrition of mycorrhizal trees. *Tree Physiology* 30, pp. 1129–1139.

Plassard, C., J. Louche, M. Ali, M. Duchemin, E. Legname, and B. Cloutier-Hurteau (2011). Diversity in phosphorus mobilisation and uptake in ectomycorrhizal fungi. *Annals of Forest Science* 68, pp. 33–43.

Plassard, C. et al. (2015). Améliorer la biodisponibilité du phosphore: comment valoriser les compétences des plantes et les mécanismes biologiques du sol. *Innovations Agronomiques* 43, pp. 115–138.

Poll, C., J. Ingwersen, M. Stemmer, M. Gerzabek, and E. Kandeler (2006). Mechanisms of solute transport affect small-scale abundance and function of soil microorganisms in the detritusphere. *European Journal of Soil Science* 57, pp. 583–595.

Puget, P. and R. Lal (2005). Soil organic carbon and nitrogen in a Mollisol in central Ohio as affected by tillage and land use. *Soil and Tillage Research* 80, pp. 201–213.

Quackenbush, J., F. Liang, I. Holt, G. Pertea, and J. Upton (2000). The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Research* 28, pp. 141–145.

Quiquampoix, H. and D. Mousain (2005). Enzymatic hydrolysis of organic phosphorus. In: *Organic phosphorus in the environment*. CAB International Wallingford, pp. 89–112.

Ragot, S., M. Kertesz, and E. Bünemann (2015). *phoD* alkaline phosphatase gene diversity in soil. *Applied and Environmental Microbiology* 81, pp. 7281–7289.

Ramaekers, L., R. Remans, I. Rao, M. Blair, and J. Vanderleyden (2010). Strategies for improving phosphorus acquisition efficiency of crop plants. *Field Crops Research* 117, pp. 169–176.

Ramette, A. (2007). Multivariate analyses in microbial ecology. *FEMS Microbiology Ecology* 62, pp. 142–160.

Reilly, T., G. Baron, F. Nano, and M. Kuhlenschmidt (1996). Characterization and sequencing of a respiratory burst-inhibiting acid phosphatase from *Francisella tularensis*. *Journal of Biological Chemistry* 271, pp. 10973–10983.

Rengel, Z. and P. Marschner (2005). Nutrient availability and management in the rhizosphere: exploiting genotypic differences. *New Phytologist* 168, pp. 305–312.

Richardson, A., P. Hadobas, and J. Hayes (2001). Extracellular secretion of *Aspergillus phytase* from *Arabidopsis* roots enables plants to obtain phosphorus from phytate. *The Plant Journal* 25, pp. 641–649.

Richardson, A., P. Hocking, R. Simpson, and T. George (2000). Plant mechanisms to optimise access to soil phosphorus. In: *Crop and Pasture Science*. CSIRO Publications, pp. 50–62.

Richardson, A., C. Pankhurst, B. Doube, V. Gupta, and P. Grace (1994). Soil microorganisms and phosphorus availability. *Soil biota: management in sustainable farming systems*, pp. 50–62.

Richardson, A. and R. Simpson (2011). Soil microorganisms mediating phosphorus availability update on microbial phosphorus. *Plant Physiology* 156, pp. 989–996.

Rivas, L., M. García-Villadangos, M. Moreno-Paz, P. Cruz-Gil, J. Gómez-Elvira, and V. Parro (2008). A 200-antibody microarray biochip for environmental monitoring: searching for universal microbial biomarkers through immunoprofiling. *Analytical Chemistry* 80, pp. 7970–7979.

Rocca, J. D. et al. (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *The ISME journal* 9, pp. 1693–1699.

Rodriguez, F., J. Lillington, S. Johnson, C. Timmel, S. Lea, and B. Berks (2014). Crystal structure of the *Bacillus subtilis*phosphodiesterase PhoD reveals an iron and calcium-containing active site. *Journal of Biological Chemistry* 289, pp. 30889–30899.

Rolfe, M. et al. (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology* 194, pp. 686–701.

Rousk, J. et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal* 4, pp. 1340–1351.

Sait, M., P. Hugenholtz, and P. Janssen (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environmental Microbiology* 4, pp. 654–666.

Sakurai, M., J. Wasaki, Y. Tomizawa, T. Shinano, and M. Osaki (2008). Analysis of bacterial communities on alkaline phosphatase genes in soil supplied with organic matter. *Soil Science and Plant Nutrition* 54, pp. 62–71.

Santos-Beneit, F. (2015). The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology* 6, a402.

Schachtman, D., R. Reid, and S. Ayling (1998). Phosphorus uptake by plants: From soil to cell. *Plant Physiology* 116, pp. 447–453.

Schefe, C. et al. (2015). 100 Years of superphosphate addition to pasture in an acid soil a current nutrient status and future management. *Soil Research* 53, pp. 662–676.

Schloss, P., B. Larget, and J. Handelsman (2004). Integration of microbial ecology and statistics: a test to compare gene libraries. *Applied and Environmental Microbiology* 70, pp. 5485–5492.

Schloss, P. et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75, pp. 7537–7541.

Schouten, S., E. Hopmans, and J. Damsté (2013). The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: a review. *Organic Geochemistry* 54, pp. 19–61.

Sebastián, M. and J. Ammerman (2009). The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *The ISME Journal* 3, pp. 563–572.

Sebastián, M. and J. Ammerman (2011). Role of the phosphatase PhoX in the phosphorus metabolism of the marine bacterium *Ruegeria pomeroyi* DSS-3. *Environmental Microbiology reports* 3, pp. 535–542.

Shen, J. (2006). Optimal estimation of parameters for a estuarine eutrophication model. *Ecological Modelling* 191, pp. 521–537.

Siciliano, S. et al. (2014). Soil fertility is associated with fungal and bacterial richness, whereas pH is associated with community composition in polar soil microbial communities. *Soil Biology and Biochemistry* 78, pp. 10–20.

Sinsabaugh, R. and D. Moorhead (1994). Resource allocation to extracellular enzyme production: a model for nitrogen and phosphorus control of litter decomposition. *Soil Biology and Biochemistry* 26, pp. 1305–1311.

Šnajdr, J., V. Valášková, V. Merhautová, T. Cajthaml, and P. Baldrian (2008). Activity and spatial distribution of lignocellulose-degrading enzymes during forest soil colonization by saprotrophic basidiomycetes. *Enzyme and Microbial Technology* 43, pp. 186–192.

Song, L. (2006). A soluble form of phosphatase in *Saccharomyces cerevisiae* capable of converting farnesyl diphosphate into E, E-farnesol. *Applied Biochemistry and Biotechnology* 128, pp. 149–157.

Von Sperber, C., H. Kries, F. Tamburini, S. M. Bernasconi, and E. Frossard (2014). The effect of phosphomonoesterases on the oxygen isotope composition of phosphate. *Geochimica et Cosmochimica Acta* 125, pp. 519–527.

Steenwerth, K., L. Jackson, F. Calderón, M. Stromberg, and K. Scow (2002). Soil microbial community composition and land use history in cultivated and grassland ecosystems of coastal California. *Soil Biology and Biochemistry* 34, pp. 1599–1611.

Stonehouse, M. et al. (2002). A novel class of microbial phosphocholine-specific phospholipases C. *Molecular Microbiology* 46, pp. 661–676.

Štursová, M., L. Žifčáková, M. B. Leigh, R. Burgess, and P. Baldrian (2012). Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiology Ecology* 80, pp. 735–746.

Su, Z., V. Olman, and Y. Xu (2007). Computational prediction of Pho regulons in *Cyanobacteria*. *BMC Genomics* 8, p. 156.

Syers, J., A. Johnston, and D. Curtin (2008). Efficiency of soil and fertilizer phosphorus use. *FAO Fertilizer and Plant Nutrition Bulletin* 18.

Takahashi, T. et al. (2014). Climatological distributions of pH, pCO_2, total CO_2, alkalinity, and $CaCO_3$ saturation in the global surface ocean, and temporal changes at selected locations. *Marine Chemistry* 164, pp. 95–125.

Tamburini, F., S. Bernasconi, A. Angert, T. Weiner, and E. Frossard (2010). A method for the analysis of the $\delta^{18}O$ of inorganic phosphate extracted from soils with HCl. *European Journal of Soil Science* 61, pp. 1025–1032.

Tamburini, F., V. Pfahler, E. Bünemann, K. Guelland, S. Bernasconi, and E. Frossard (2012). Oxygen isotopes unravel the role of microorganisms in phosphate cycling in soils. *Environmental Science and Technology* 46, pp. 5956–5962.

Tamburini, F., V. Pfahler, C. von Sperber, E. Frossard, and S. Bernasconi (2014). Oxygen isotopes for unraveling phosphorus transformations in the soil–plant system: A review. *Soil Science Society of America Journal* 78, pp. 38–46.

Tan, H. et al. (2013). Long-term phosphorus fertilisation increased the diversity of the total bacterial community and the *phoD* phosphorus mineraliser group in pasture soils. *Biology and Fertility of Soils* 49, pp. 661–672.

Tarafdar, J., R. Yadav, and S. Meena (2001). Comparative efficiency of acid phosphatase originated from plant and fungal sources. *Journal of Plant Nutrition and Soil Science* 164, pp. 279–282.

Tarafdar, J., R. Yadav, and R. Niwas (2002). Relative efficiency of fungal intra-and extracellular phosphatases and phytase. *Journal of Plant Nutrition and Soil Science* 165, pp. 17–19.

Toh-e, A., Y. Ueda, S.-T. Kakimoto, and Y. Oshima (1973). Isolation and characterization of acid phosphatase mutants in *Saccharomyces cerevisiae*. *Journal of Bacteriology* 113, pp. 727–738.

Touati, E. and A. Danchin (1987). The structure of the promoter and amino terminal region of the pH 2.5 acid phosphatase structural gene (*appA*) of *E. coli*: a negative control of transcription mediated by cyclic AMP. *Biochimie* 69, pp. 215–221.

Turner, B., B. Cade-Menun, and D. Westermann (2003). Organic phosphorus composition and potential bioavailability in semi-arid arable soils of the western United States. *Soil Science Society of America Journal* 67, pp. 1168–1179.

Van Bogelen, R., E. Olson, B. Wanner, and F . Neidhardt (1996). Global analysis of proteins synthesized during phosphorus restriction in *Escherichia coli*. *Journal of Bacteriology* 178, pp. 4344–4366.

Vance, E., P. Brookes, and D. Jenkinson (1987). An extraction method for measuring soil microbial biomass C. *Soil biology and Biochemistry* 19, pp. 703–707.

Vershinina, O. and L. Znamenskaya (2002). The Pho regulons of bacteria. *Microbiology* 71, pp. 497–511.

Vial, L., A. Chapalain, M.-C. Groleau, and E. Déziel (2011). The various lifestyles of the *Burkholderia cepacia* complex species: a tribute to adaptation. *Environmental Microbiology* 13, pp. 1–12.

Vidal, M., C. M. Duarte, S. Agustí, J. M. Gasol, and D. Vaqué (2003). Alkaline phosphatase activities in the central Atlantic Ocean indicate large areas with phosphorus deficiency.

Vitousek, P., S. Porder, B. Houlton, and O. Chadwick (2010). Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions. *Ecological Applications* 20, pp. 5–15.

Vogeler, I., J. Rogasik, U. Funder, K. Panten, and E. Schnug (2009). Effect of tillage systems and P-fertilization on soil physical and chemical properties, crop yield and nutrient uptake. *Soil and Tillage Research* 103, pp. 137–143.

Wakelin, S., L. Macdonald, S. Rogers, A. L. Gregg, T. Bolger, and J. A. Baldock (2008). Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. *Soil Biology and Biochemistry* 40, pp. 803–813.

Walker, T. and J. K. Syers (1976). The fate of phosphorus during pedogenesis. *Geoderma* 15, pp. 1–19.

Wallenstein, M., S. McNulty, I. Fernandez, J. Boggs, and W. Schlesinger (2006). Nitrogen fertilization decreases forest soil fungal and bacterial biomass in three long-term experiments. *Forest Ecology and Management* 222, pp. 459–468.

Walters, W., J. Caporaso, C. Lauber, D. Berg-Lyons, N. Fierer, and R. Knight (2011). PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics* 27, pp. 1159–1161.

Wang, J., K. Stieglitz, and E. Kantrowitz (2005). Metal specificity is correlated with two crucial active site residues in *Escherichia coli* alkaline phosphatase. *Biochemistry* 44, pp. 8378–8386.

Wang, Y., P. Marschner, and F. Zhang (2012a). Phosphorus pools and other soil properties in the rhizosphere of wheat and legumes growing in three soils in monoculture or as a mixture of wheat and legume. *Plant and Soil* 354, pp. 283–298.

Wang, Y., F. Zhang, and P. Marschner (2012b). Soil pH is the main factor influencing growth and rhizosphere properties of wheat following different pre-crops. *Plant and Soil* 360, pp. 271–286.

Ward, A. and N. Bora (2008). The *Actinobacteria*. In: *Practical handbook of microbiology*. 2nd edn. CRC Press, London, pp. 375–444.

Weaver, R. et al. (1994). *Methods of soil analysis: part 2, microbiological and biochemical properties*. Soil Science Society of America Madison, WI.

Weintraub, M. and J. Schimel (2005). The seasonal dynamics of amino acids and other nutrients in Alaskan Arctic tundra soils. *Biogeochemistry* 73, pp. 359–380.

Wende, A., P. Johansson, R. Vollrath, M. Dyall-Smith, D. Oesterhelt, and M. Grininger (2010). Structural and biochemical characterization of a halophilic archaeal alkaline phosphatase. *Journal of Molecular Biology* 400, pp. 52–62.

Westheimer, F. (1987). Why nature chose phosphates. *Science* 235, pp. 1173–1178.

Wharton, C. (1969). *The Green Revolution: cornucopia or Pandora's box?* Council on Foreign Relations, Incorporated.

White, T., T. Bruns, S. Lee, and J. Taylor (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications* 18, pp. 315–322.

Wilson, M. et al. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506, pp. 58–62.

Winsley, T., J. Van Dorst, M. Brown, and B. Ferrari (2012). Capturing greater 16S rRNA gene sequence diversity within the domain bacteria. *Applied and Environmental Microbiology* 78, pp. 5938–5941.

Wu, J.-R. et al. (2007). Cloning of the gene and characterization of the enzymatic properties of the monomeric alkaline phosphatase (PhoX) from *Pasteurella multocida* strain X-73. *FEMS Microbiology Letters* 267, pp. 113–120.

Yamane, K. and B. Maruo (1978). Purification and characterization of extracellular soluble and membrane-bound insoluble alkaline phosphatases possessing phosphodiesterase activities in *Bacillus subtilis*. *Journal of Bacteriology* 134, pp. 100–107.

Yang, H. and M. Roberts (2004). Expression and characterization of a heterodimer of *Streptomyces chromofuscus* phospholipase D. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1703, pp. 43–51.

Yang, M., Z. Wang, M. MacPherson, J. Dow, and K. Kaiser (2000). A novel *Drosophila* alkaline phosphatase specific to the ellipsoid body of the adult brain and the lower Malpighian (renal) tubule. *Genetics* 154, pp. 285–297.

Yang, X. and W. Post (2011). Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using Hedley fractionation method. *Biogeosciences* 8, pp. 2907–2916.

Yong, S. et al. (2014). A complex iron-calcium cofactor catalyzing phosphotransfer chemistry. *Science* 345, pp. 1170–1173.

Yu, H. et al. (2007). Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation* 18, pp. 793–802.

Zaheer, R., R. Morton, M. Proudfoot, A. Yakunin, and T. Finan (2009). Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. *Environmental Microbiology* 11, pp. 1572–1587.

Zappa, S., J.-L. Rolland, D. Flament, Y. Gueguen, J. Boudrant, and J. Dietrich (2001). Characterization of a highly thermostable alkaline phosphatase from the euryarchaeon *Pyrococcus abyssi*. *Applied and Environmental Microbiology* 67, pp. 4504–4511.

Zhang, L. and G. Xie (2007). Diversity and distribution of Burkholderia cepacia complex in the rhizosphere of rice and maize. *FEMS Microbiology Letters* 266, pp. 231–235.

Zhong, W. and Z. Cai (2007). Long-term effects of inorganic fertilizers on microbial biomass and community functional diversity in a paddy soil derived from quaternary red clay. *Applied Soil Ecology* 36, pp. 84–91.

Zimmerman, A., A. Martiny, and S. Allison (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *The ISME Journal* 7, pp. 1187–1199.

Zimmermann, P. et al. (2003). Engineering the root–soil interface via targeted expression of a synthetic phytase gene in trichoblasts. *Plant Biotechnology Journal* 1, pp. 353–360.

ACKNOWLEDGEMENTS

My time as a doctoral student has been a wonderful learning experience thanks to the help and support of many people, also those not mentioned explicitly.

My deepest thanks go to:

Prof. Dr. Emmanuel Frossard, for giving me the opportunity to work in the Plant Nutrition Group, a highly professional environment with passionate and friendly people. He always challenged and inspired me to dig deeper and broaden my knowledge. He gave me constructive and practical advice and personal support. I will never forget his authentic and powerful laugh!

PD Dr. Else Bünemann for her passion for soil sciences and phosphorus cycling, and for trusting this project to me. She gave me excellent supervision and valuable advice throughout each step of this work. Her guidance, advice, patience and stimulating conversations were essential to the completion of this project. Thanks for forgiving my spelling mistakes and always greeting me with a smile!

Prof. Dr. Michael Kertesz for welcoming me in his group at the University of Sydney and guiding me throughout the design and validation of the primers. His expertise in microbial ecology was always insightful. Thanks for helping me in the field and showing me the beauty of Downunder!

Dr. Claude Plassard for reviewing my work as an external examiner and giving me constructive advice to improve my thesis.

Dr. Olivier Huguenin for his valuable comments and discussion on our study in Watt, and for plant data.

Dr. Laurie Schönholzer for analyzing a plethora of samples and for giving heaps of tips and advice. Thanks for being so enthusiastic and energetic!

Klaus Jarosch for the great discussions, support and inspiration during our doctoral studies. I am still hoping for a joined research project! Thanks for being ready for a laugh anytime of the day!

Dr. Éva Mészáros for her efficient and accurate work which generated the majority of the data in Chapter 4. I am grateful for the constructive discussion on molecular biology and data analysis and your help with the design . Thanks for your kindness and for always being in a wonderful mood!

Monika Mascai for her help in the fieldwork of Chapter 4 and the plant analysis in Chapter 5.

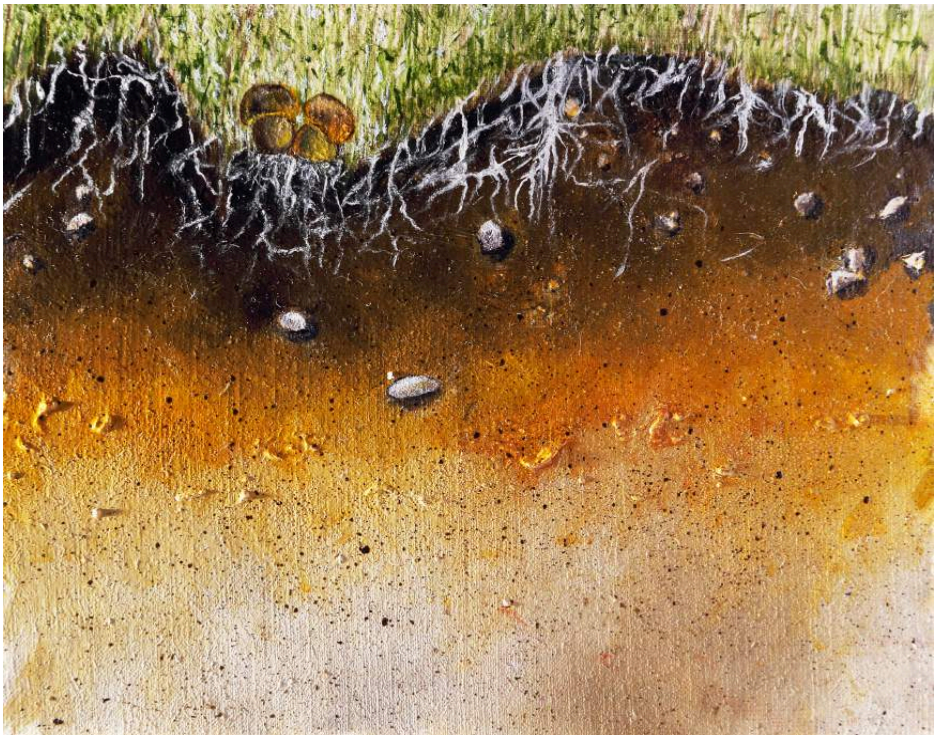
The group of Plant Nutrition at ETH for the amazing work atmosphere, coffee breaks, laughter and support. Thanks to Dr. Federica Tamburini, Dr. Verena Pfahler, Dr. Christian Von Sperber, Dr. Simone Nanzer, Dr. Kathrin Annaheim, Dr. Hanner Gamper, Dr. Astrid Oberson, Dr. Chiara Psitocchi, Dr. Timothy Mclaren, Dr. Stephanie Cheesman, Dr. Seher Bahar Aciksöz Özden, Dr. Chiara Pistocchi, Dr. Adibine Traore, Oliver Zemek, Christiane Gujan, Julian Helfenstein, Benjamin Costerousse, Gregor Meyer, Thilo Dürr-Auster, Samuel Mathu Ndungu, Ravinda Lakshan Senanayake, Renata Slavikova, Pierre Stevenel, Matthias Wiggenhauser and Salvatore Accardo.

The group of Soil Microbiology at the University of Sydney for welcoming and supporting me. Dr. Neil Wilson for his advice on molecular biology, humor and patience. Dr. Paola E. Corneo for her advice, support and amazing personality, and for her friendship. Dr. Katarzyna Safianowics, Dr. Mirja Guldner, Nanako Horimoto and Claudia Carrasco for their kindness and support.

Last, but absolutely not least, I give my thanks and credit to my parents and family for their love, encouragement and support. They let me define my own path and supported

every decision I made. My mother for her fabulous paintings that bring beauty and colors to my research. Pratik for his unconditional support and patience through the peaks and valleys of this doctoral studies experience.

This work was made possible by the funding from the Swiss National Funds (project number 140900), the University of Sydney and ETH Zürich, and the field trials of the University of Sydney and Agroscope.



Artist: Anne Ragot

„The role of the infinitely small is infinitely large."

Louis Pasteur