

## RESEARCH ARTICLE

# Abundance and phylogenetic distribution of eight key enzymes of the phosphorus biogeochemical cycle in grassland soils

Silvia Garaycochea<sup>1</sup>  | Nora Adriana Altier<sup>1</sup>  | Carolina Leoni<sup>1</sup>  |  
Andrew L. Neal<sup>2</sup>  | Héctor Romero<sup>3</sup>

<sup>1</sup>Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental INIA Las Brujas, Canelones, Uruguay

<sup>2</sup>Net-Zero and Resilient Farming, Rothamsted Research, North Wyke, Okehampton, UK

<sup>3</sup>Laboratorio de Organización y Evolución del Genoma/Genómica Evolutiva, Departamento de Ecología y Evolución, Facultad de Ciencias/CURE, Universidad de la República, Maldonado, Uruguay

## Correspondence

Héctor Romero, Laboratorio de Organización y Evolución del Genoma/Genómica Evolutiva, Departamento de Ecología y Evolución, Facultad de Ciencias/CURE, Universidad de la República, Maldonado, Uruguay.  
Email: [eletor@fcien.edu.uy](mailto:eletor@fcien.edu.uy)

Silvia Garaycochea, Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental INIA Las Brujas, Ruta 48 Km 10, Canelones, 90200, Uruguay.  
Email: [sgaraycochea@inia.org.uy](mailto:sgaraycochea@inia.org.uy)

## Funding information

Agencia Nacional de Investigación e Innovación, Grant/Award Number: POS NAC 2015 1 110075; Instituto Nacional de Investigación Agropecuaria, INIA, Grant/Award Numbers: Project SA47, SA 26, SA 24

## Abstract

Grassland biomes provide valuable ecosystem services, including nutrient cycling. Organic phosphorus (Po) represents more than half of the total P in soils. Soil microorganisms release organic P through enzymatic processes, with alkaline phosphatases, acid phosphatases and phytases being the key P enzymes involved in the cycling of organic P. This study analysed 74 soil metagenomes from 17 different grassland biomes worldwide to evaluate the distribution and abundance of eight key P enzymes (PhoD, PhoX, PhoA, Nsap-A, Nsap-B, Nsap-C, BPP and CPhy) and their relationship with environmental factors. Our analyses showed that alkaline phosphatase phoD was the dataset's most abundant P-enzyme encoding genes, with a wide phylogenetic distribution. Followed by the acid phosphatases Nsap-A and Nsap-C showed similar abundance but a different distribution in their respective phylogenetic trees. Multivariate analyses revealed that pH,  $T_{max}$ , SOC and soil moisture were associated with the abundance and diversity of all genes studied. PhoD and phoX genes strongly correlated with SOC and clay, and the phoX gene was more common in soils with low to medium SOC and neutral pH. In particular, P-enzyme genes tended to respond in a positively correlated manner among them, suggesting a complex relationship of abundance and diversity among them.

## INTRODUCTION

Grasslands are one of the most numerous and widely distributed biomes on the Earth's surface. Factors defining grassland biomes are climatic conditions, grazing and fire (White et al., 2000; Zhou et al., 2017). They develop in arid and semi-arid areas, with seasonal cold and dry periods, and high rates of evapotranspiration (Barnett & Facey, 2016; Knapp et al., 2002; Lenhart et al., 2015). The plant community is dominated by

grasses and grass-like species, as well as other shrubby species with different lifestyles. Plant community assemblages depend largely on climatic variables. Most of the grassland biomass above-ground, together with the low rates of decomposition, generates significant accumulations of organic matter in soil profiles (Blair et al., 2014). Grasslands also provide several key ecosystem services, such as food, fibre and forage production, water and nutrient cycling, and erosion control. Grassland biomes are habitats for a high diversity of

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Environmental Microbiology Reports* published by Applied Microbiology International and John Wiley & Sons Ltd.

plants, animals and microorganisms (Blair et al., 2014; Le Roux et al., 2011).

Nutrient cycling, one of the main ecosystem services provided by grasslands, can be defined as the cycling of elements carbon (C), nitrogen (N) and phosphorus (P) between different pools (Dubeux et al., 2007). Soils have low P availability as a result of high reactivity with calcium (Ca), iron (Fe), or aluminium (Al) ions, forming insoluble complexes (Achat et al., 2016). Soil P is present in two fractions, organic (Po) and inorganic phosphates (Pi) whose proportions between soils vary depending mainly on the geological material, pH, temperature and organic matter contributions (Gaiero et al., 2020; Zhou et al., 2017). On average, the organic fraction accounts for over half of total soil P and is a valuable reservoir that could be partially mobilized by microorganisms (Condrón et al., 2005; George et al., 2018; Haygarth et al., 2013). The more abundant organic P forms in soils are inositol phosphate, phospholipids, nucleic acids and teichoic acid (Condrón et al., 2005; Gyaneshwar et al., 2002). Inositol phosphate (commonly called phytic acid) can account for up to 80% of total organic P (Gerke, 2015; Quiquampoix & Mousain, 2005). Phytic acid reacts with ions present in the soil forming stable and insoluble complexes and so tends to accumulate in natural grasslands soils. On the other hand, phospholipids and nucleic acids are both labile and readily accessible to soil organisms (Gerke, 2015).

The Po mineralization is strongly influenced by several factors, including soil pH, total N, precipitation and temperature, and is mediated by various enzymes with phosphatase activity. These enzymes, which are involved in different stages of the P cycle, are also influenced by such environmental factors (Margalef et al., 2017). Po-cycle genes can be divided into three groups: Po mineralization (e.g. *phoD*, *phy*, *phoC*), transporter genes (e.g. *pstS*, *ugpQ*), and P starvation regulation genes (e.g. *phoB*, *phoR*) (Bergkemper et al., 2016; Oliverio et al., 2020; Zeng et al., 2022). The Po mineralization genes encode enzymes capable of releasing P from organic phosphate esters (henceforth P-enzymes). The alkaline phosphatases and non-specific acid phosphatases (Nsap) catalyse the hydrolysis between carbon and phosphorus in organic phosphate esters. The third group, the phytases, specifically release Pi from phytic acid (Bergkemper et al., 2016; Gaiero et al., 2020; Huang et al., 2009; Jorquera et al., 2008; Morrison et al., 2016; Rossolini et al., 1998). The two-component regulatory system (PhoBR) encoded by *phoBR*, called the Pho regulon, regulates the transcription of P-enzyme genes under low Pi conditions (Lidbury et al., 2017; Park et al., 2022; Santos-Beneit et al., 2015). Alkaline phosphatases are produced by a broad range of bacteria, archaea and fungi, which play an important role in microbial P turnover (Li et al., 2021). *PhoD*, *PhoX* and

*PhoA* are three different types of alkaline phosphatases, with *PhoD* being the most abundant and ubiquitous (Ragot et al., 2015). Both *PhoD* and *PhoX* were identified as Ca<sup>2+</sup>-dependent extracellular enzymes and *PhoA* as a Zn<sup>2+</sup>-dependent intracellular enzyme (Neal et al., 2018). Alkaline phosphatases show a broad substrate specificity and high catalytic efficiency (Cai et al., 2021; Rodriguez et al., 2014). These characteristics enable microorganisms harbouring these genes to use alternative P sources under P-limited conditions, conferring them an advantage over the plants (Li et al., 2021).

Acid phosphatases are another group of enzymes distributed widely among microorganisms and plants. They are divided into three groups, Nsap class A, Nsap class B and Nsap class C, none of which exhibit strong substrate specificity, hence their names (Thaller et al., 1998). These enzymes are mostly produced by microorganisms and are mostly active in acid soils (Gaiero et al., 2018). To expand the knowledge of these enzymes, metagenomic studies have been carried out to understand how they vary in abundance and diversity in different environments (Bergkemper et al., 2016; Neal et al., 2018). Neal et al. (2018) showed that Nsap class C, a putative extracellular enzyme, was predominant in acid soils under P-limiting conditions compared with Nsap class A a putative intracellular or periplasmic enzyme. These enzyme groups have been observed to have higher activity and gene abundance in the rhizosphere than in the bulk soil (Fraser et al., 2017; Spohn & Kuzyakov, 2013).

Phytases are produced by bacteria, fungi, plants and animals able to catalyse the mineralization of organic P from phytate to inorganic P (Ariza et al., 2013; Jorquera et al., 2008; Tu et al., 2011). Phytase families, more common in microorganisms, are the beta-propeller phytase (BPP), protein tyrosine phosphatase-like cysteine phytase (CPhy) and histidine acid phytase (HAPhy) (Lim et al., 2007). The main differences between the phytase families are structural, mainly related to differences in the active site which determines which phosphate group of the phytate is dephosphorylated, and co-factor requirements. Despite this, all phytases can release the six phosphate molecules contained in the phytate (Misset, 2002). Phytases exhibit different pH and temperature optima in the laboratory (Caffaro et al., 2020) and also are dependent on the soil microorganisms species (Amadou et al., 2021). Moreover, enzymatic activity is affected by soil type, texture and mineralogy by varying the ability to retain an active enzyme (Azeem et al., 2015; Rao et al., 1994; Tang et al., 2006).

Soil microorganisms play an important role in the soil P cycle, mediating P release for plants and other living soil organisms (Awasthi et al., 2011; Richardson & Simpson, 2011). Several prokaryotic phyla have been associated with soil Po mineralization

*Acidobacteria*, *Actinobacteria*, *Firmicutes* and *Proteobacteria* (Amadou et al., 2021). These Po mineralizing phyla contain a repertoire of genes that allow them to obtain Pi from organic compounds using different strategies. Forest soil study showed *Actinobacteria* and *Proteobacteria* played a dominant role in oxidative phosphorylation, whereas *Firmicutes* contributed to substrate phosphorylation (Ma et al., 2021). The alkaline phosphatase encoded by the *phoD* gene was primarily found in bacteria and was spread across 20 bacterial phyla (Ragot et al., 2015). Grassland microbiome studies showed *Actinobacteria*, *Planctomycetes* and *Proteobacteria* were the dominant bacterial phyla carrying the *phoD* gene, representing over 80% of all sequences (Graça et al., 2021). The *Streptomyces* genomes harbour alkaline phosphatases encoded by *phoA* and *phoD* genes and acid phosphatase class A coding gene (*phoC*) (Tian et al., 2021). Finally, the *Streptococcus* genus has been associated with phytase production and mineralization of phosphate (de Lacerda et al., 2016).

Grasslands are one of the five most important biomes on Earth due to the biodiversity they harbour and their economic importance. This makes it necessary to have a deeper understanding of its functions and dynamics for its preservation. This study aimed, through a global scale analysis of metagenomic data, to assess how eight key prokaryotic P-enzymes involved in P cycling vary in their abundance and diversity in grassland biomes, how are they related between them, how they interact with the general functional profiles, and how is this related to environmental variables. We hypothesized a certain association between the P-enzyme coding genes, and that the different soil properties and climate variables of grassland would affect the profiles of these genes. We then attempted to identify which variables could be drivers of the observed patterns.

## EXPERIMENTAL PROCEDURES

### Data collection

A total of 376 geo-referenced metagenome samples from 17 projects deposited with MG-RAST were selected through the TerrestrialMetagenomeDB (<https://webapp.ufz.de/tmdb/>) applying the following filters: Source DB: MG-RAST; seq\_technology: Illumina; material: soil; Biome: grasslands, temperate grasslands, savanna and shrubland to assemble the grassland soil metagenomes samples set (Figure S1). All metagenomes included in the dataset were from topsoil samples (depth 10–15 cm). The set of environmental variables was assembled, including soil properties and climatic variables for each sample based on its geographic location. Soil type and physicochemical

properties were obtained from SoilGrid 250 m 2.0 – ISRIC World Soil Information. The following properties were included Bulk Density (BD;  $\text{cg cm}^{-3}$ ), Clay ( $\text{g kg}^{-1}$ ), Sand ( $\text{g kg}^{-1}$ ), Silt ( $\text{g kg}^{-1}$ ), Cation Exchange Capacity at pH 7 (CEC;  $\text{mmol(c) kg}^{-1}$ ), Total Nitrogen (N;  $\text{dg kg}^{-1}$ ), Soil Organic Carbon (SOC;  $\text{dg kg}^{-1}$ ), pH ( $\text{water} \times 10$ ). The estimated organic available P (Pav) was estimated based on SOC and N content following a model proposed by Tian et al. (2010) who proposed a C:N:P ratio of 134:9:1 for organic-rich topsoil, we estimated P content in relation to C:P and N:P ratios and took as P value the average between them. Climate variables were obtained from TerraClimate (<https://www.climatologylab.org/terraclimate.html>), including maximum temperature, ( $T_{\text{max}}$ ; °C), Precipitation (ppt; mm), actual evapotranspiration (aet; mm), soil moisture (moisture; mm) and runoff (q; mm) (Table S1). Hereafter they are called environment variables. The collinearity analysis on the environmental variables set was performed with R-base (R core Team 2022), We included variables with  $r \leq 0.5$  and meaningful to the study.

The functional annotation based on MG-RAST subsystems level 2 of the 376 selected metagenomes (Table S2) was obtained from the MG-RAST repository (Meyer et al. 2008).

The set of predicted proteins in each metagenome was obtained through the RESTful API of MG-RAST (Wilke et al., 2015). Protein sequences were downloaded using a matR version 0.9.1 package R (Braithwaite & Keegan, 2018).

The set of 376 samples showed imbalances because of the overrepresentation of the same sites, particularly from the northern hemisphere (much more studied) compared with the southern hemisphere. To minimize this bias, subsequent analyses were performed on a balanced reduced subset of 74 grasslands soil metagenomes. This subset included a maximum of three samples per MG-RAST project with the same geo-reference. In addition, soil metagenome data from two Uruguayan sites were generated for this study (Table S3). In the subset, we excluded the samples under high-impact treatments (e.g. fertilization, tillage, etc.). All analyses were performed on this reduced subset of 74 samples from 17 MG-RAST projects (Table S3).

Soil metagenomic sequencing from Uruguay (projects mgp91922 and mgp93346) was carried out on a HiSeq Illumina platform, (Service CD Genomics, NY; pair-end read 150 bp). Raw sequence quality was analysed with FastQC software version 0.11.2. Assembly and functional annotation were performed on the MG-RAST repository. Raw sequence data are publicly available on the MG-RAST repository. Functional annotation based on MG-RAST subsystems level 2 of the 376 selected metagenomes was obtained from the MG-RAST repository (Meyer et al. 2008). The set of

predicted proteins of each metagenome was obtained through the RESTful API of MG-RAST (Wilke et al., 2015). Protein sequences were downloaded using matR version 0.9.1 package R (Braithwaite & Keegan, 2018).

## P-enzyme gene identification and phylogenetic analyses

The reference databases of the P-enzyme used in this work were built by Neal et al. (2017). The P-enzymes included are listed in Table 1. It is important to note that the use of any reference database introduces a certain bias in the search space.

Protein sequence alignments of the respective reference database were performed using MAFFT version 7.4.60 (Kato et al., 2002) under default parameters. Reference protein phylograms were inferred with IQTree 2 version 1.6.12 (Minh et al., 2020) and the evolutionary models were evaluated with RAXML-NG (Kozlov et al., 2019). Phylograms were plotted with iTOL (Interactive Tree of Life; Letunic & Bork, 2007).

To determine the abundance and diversity of the P-enzymes in the metagenomes, we queried each metagenomic sample against each P-enzyme reference database. First, the whole predicted protein set of each metagenomic sample was queried against each P-enzyme reference database using HMMER version 3.3.1 (<http://hmmer.org>) keeping hits with an e-value below  $1e^{-5}$ . Then, these sequences were aligned to the correspondent reference database alignment using MAFFT—add sequence option and default parameters.

Maximum likelihood-based phylogenetic placement of metagenome-derived protein sequences on the appropriate P-enzyme reference phylogenetic tree was performed with EPA-ng (Barbera et al., 2019). Edge-PCA ordination and Kantorovich-Rubinstein (KR) distance metrics (Evans & Matsen, 2012; Matsen & Evans, 2013) were computed on these results. The edge-PCA and KR distances were

performed using gappa (Czech et al., 2020), and tree and domain composition diagrams were drawn using Archaeopteryx (<https://sites.google.com/site/cmzmas/ek/home/software/forester>).

## Statistical analyses

Canonical analysis of principal coordinates (CAP) (Anderson & Willis, 2003) implemented in Vegan R Package version 2.6.2 (Oksanen et al., 2019) was performed based upon Mahalanobis distance to calculate the relationship between the metagenomes functional profiles (subsystems level 2) and environmental variables. The significance of the model parameters was determined with permutational multivariate analysis of variance (PERMANOVA) with 999 permutations.

The protein/function count matrix (level 4 in the MG-RAST nomenclature), including the eight P-enzymes, for the 74 selected metagenomes was normalized with CPM and TMM methods using the edgeR package (Robinson et al., 2010). This data was used to perform the direct correlations of P-enzymes with environmental variables.

Canonical analysis of principal coordinates (CAP) (Anderson & Willis, 2003) implemented in the Vegan R Package version 2.6.2 (Oksanen et al., 2019) was used to evaluate the relationship between the abundance and diversity of P-specific functions with the environmental variables. CAP analysis associating P-enzyme abundance with environmental variables was performed using Mahalanobis distance. When appropriate, each P-enzyme abundance in each sample was normalized in relation to the sequencing coverage of each P-enzyme. The significance of the model parameters was determined with permutational multivariate analysis of variance (PERMANOVA) with 999 permutations.

The KR distance of each P-enzyme calculated as mentioned above was used to perform the distance-based CAP analyses between the abundance and diversity of each P-enzyme and environmental variables. The significance of the model parameters was

TABLE 1 List of P-enzymes included in the analyses.

P-enzyme	Gene	Predicted cellular localization	Number of protein sequences in the reference database
PhoA	<i>phoA</i>	Periplasmic/Cytoplasmic	293
PhoD	<i>phoD</i>	Outer membrane/extracellular	833
PhoX	<i>phoX</i>	Outer membrane/extracellular	424
Nsap class A (Nsap-A)	<i>phoC</i>	Periplasmic/Cytoplasmic	750
Nsap class B (Nsap-B)	<i>aphA</i>	Periplasmic/Cytoplasmic	388
Nsap class C (Nsap-C)	<i>olpA</i>	Outer membrane/extracellular	1123
$\beta$ -propeller phytase (BPP)	<i>phyL</i> , <i>phyS</i>	Outer membrane/extracellular	108
Cysteine phytase (Cphy)	<i>phyA</i>	Outer membrane/extracellular	122





also determined with PERMANOVA based on 999 permutations. Graphics were produced with the R package ggplot2 (Wickham, 2016). All basic statistical procedures were performed using R-base (R core Team 2022). All taxonomy names cited are mentioned in italics and agree with Thines et al., 2020.

## RESULTS

### Metagenome functional profiles and environmental variables

First, we wanted to generate a general perspective of grassland functional landscapes and their relationship with environmental variables. To this aim, we performed a Canonical Analysis of Principal Coordinates (CAP) on a set of 74 grassland soil metagenomes (a reduced data set to correct for imbalances in the sample number per site, see methods). We used as input 168 functional processes (level 2 of the subsystems annotation from MG-RAST, Table S4) and their corresponding environmental variables (Table S3). The constrained model was significant ( $p = 0.001$ ) and explained 24.8% of the total variance observed in the data set. Significant associations ( $r > |0.20|$ ,  $p < 0.01$ ) between the distribution of metagenomes and nine environmental variables were identified. CAP1 axis was correlated with pH ( $r = -0.743$ ), bulk density (BD,  $r = -0.521$ ), soil organic carbon (SOC,  $r = 0.564$ ) and soil moisture ( $r = 0.536$ ). This axis separated samples from low pH soils with average values of 5.65 (e.g. mgp9904, mgp5588, mgp91922 and mgp93346) from those with neutral pH (mgp13948 among others). CAP2 axis was mainly associated with pH ( $r = -0.486$ ), SOC ( $r = -0.220$ ),  $T_{\max}$  ( $r = 0.664$ ), runoff ( $q$ ,  $r = 0.490$ ), soil moisture ( $r = 0.476$ ) and Clay ( $r = 0.231$ ). Extreme values of the CAP2 axis corresponded to mgp10450 and mgp10451 (both from Brazil) which were associated with the highest  $T_{\max}$  (26°C), soil moisture (115.5 mm) and precipitation (ppt) (129 mm) values of the set (Table 3A and Figure S2). To validate the subsampling (74 vs. 376 sample set), we performed CAP analysis in the larger set and examined the correlation between the axes of both analyses. We observed a high positive correlation between the correspondent first and second axes (correlation values  $> 0.60$ ).

### Analyses on the abundance of P-enzymes coding genes

We interrogated the predicted protein set of each metagenome against the reference database of PhoD, PhoX, PhoA, Nsap-A, Nsap-B, Nsap-C, BPP and CPhy enzymes to obtain the abundance and phylogenetic

distribution of P-enzyme coding genes. Inferred protein relative abundance in each soil metagenome is shown in Table S7 and phylogenetic placements are in Figures 1 and S3.

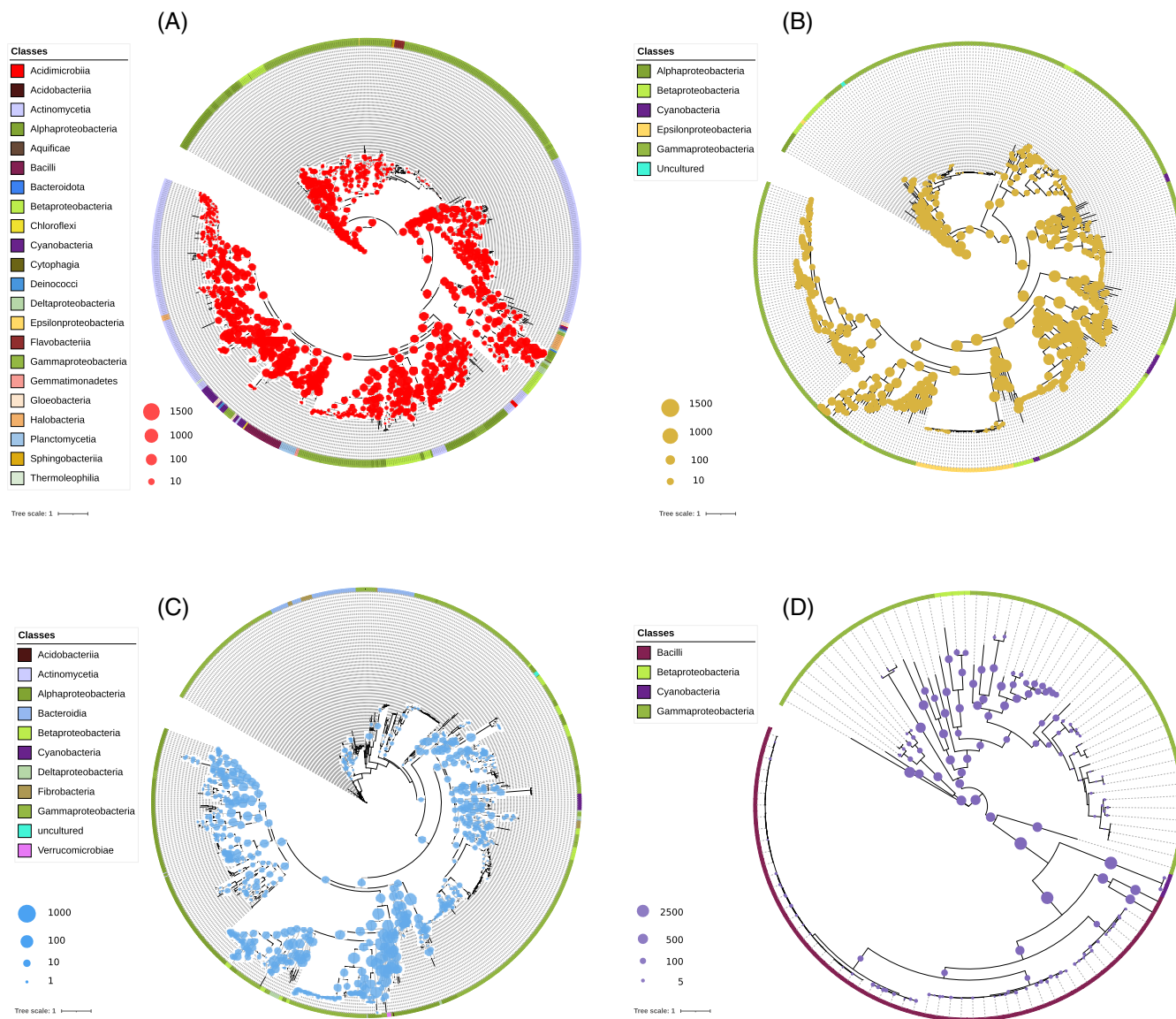
The alkaline phosphatase genes were the most abundant in the dataset, eight times higher than the acid phosphatase genes and 58 times higher than the phytase genes, independently of the soil properties (Table 2). We also observed differences in the abundance and phylogenetic distribution within each group of P-enzyme genes. The alkaline phosphatase gene *phoD* showed an abundance of five times higher compared with *phoX* and 20 times higher compared with *phoA*. Both genes, *phoD*, and *phoX* had broad phylogenetic distributions and no clear dominant phylotypes (Figures 1 and S3), contrary to the limited phylogenetic distribution observed in *phoA* (Figure S3 and Table S5).

Genes encoding Nsap-A and Nsap-C were the most abundant of the acid phosphatases, with similar abundances (Nsap-C coding gene was 1.3 times higher than Nsap-A one) (Table 2, Table S5), but a different distribution in their corresponding phylogenetic trees. Whilst Nsap-A coding gene showed a broad distribution within its phylogeny (Figure 1c), Nsap-C one was concentrated in the main branches of *Gammaproteobacteria*, *Flavobacteria* and *Sphingobacteria* classes (Figure S3). On the other side, Nsap-B had a low abundance and only *Gammaproteobacteria* variants were found (Figure S3 and Table S5).

BPP coding gene (*phyL* and *phyS*) was the most abundant of the phytases genes and presented a phylogenetic distribution mainly restricted to the *Proteobacteria* phylum (e.g. *Pseudomonas*, *Alteromonas* and *Acinetobacter*) (Figure 1d and Table 2). The CPhy coding gene (*phyA*), with lower abundance, was distributed within *Betaproteobacteria*, *Gammaproteobacteria* and some classes of the *Firmicutes* phylum (Figure S3 and Table 2).

First, we performed simple correlation analyses between normalized genes encoding P-enzyme abundance (by CPM and TMM methods, obtaining equivalent results) and environmental variables showed that *phoD*, *phoX* and *phyL* and *phyS* (BPP) coding gene had a significant correlation ( $p < 0.001$ ) with pH, actual evapotranspiration (aet), precipitation (ppt), runoff ( $q$ ) and soil moisture. In addition, we observed that *phoD* showed significant correlations ( $p < 0.001$ ) with SOC and estimated organic available P (Pav). Nsap-C coding gene (*olpA*) showed a significant correlation with aet,  $q$ , ppt and moisture (Table S6). We then move forward to multivariate analyses.

We used CAP analysis to explore the relationship between P-enzyme coding genes normalized abundance and environmental variables (Figure S4). The constrained model based on Mahalanobis distance explained 36.4% of the variance within the data set



**FIGURE 1** Phylogenetic placements of the predicted proteins of each metagenome with respect to the reference bases of each enzyme: (A) PhoD, (B) PhoX, (C) Nsap-A and (D) BPP. The size of the circle representing placements is proportional to the abundance. Maximum likelihood-based phylogenetic placement of metagenome-derived protein sequences was performed with EPA-ng and a tree was drawn with iTOL. The circle sizes represent the number of hits per node. The outer circle shows bacterial classes included in the reference trees.

**TABLE 2** Median relative abundance of each P-enzyme.

P-enzyme	PhoD	PhoX	PhoA	Nsap-A	Nsap-B	Nsap-C	BPP	Cphy
Gene	<i>phoD</i>	<i>phoX</i>	<i>phoA</i>	<i>phoC</i>	<i>aphA</i>	<i>olpA</i>	<i>phyL, phyS</i>	<i>phyA</i>
Median relative abundance (No. of hits)	468.6	96.8	23.3	29.7	0.55	39.9	10.3	0

( $p = 0.001$ ). We identified the alkaline phosphatase genes *phoD*, *phoX* and *phoA* were mainly responsible for the explained variance. CAP1 axis explained 8.9% of the variance ( $p = 0.001$ ) and was associated with pH ( $r = -0.70$ ), BD ( $r = -0.48$ ), Sand ( $r = -0.40$ ), ppt ( $r = 0.74$ ), aet ( $r = 0.67$ ), SOC ( $r = 0.52$ ), Pav ( $r = 0.52$ ) and Silt ( $r = 0.49$ ). We observed that this axis separated samples from metagenomes of clay

soils with low pH (5.0–6.8) and high SOC values (Androsols, Cambisols, Ferrasols, Fluvisols, Kastanozem/Luvisol, Luvisol/Kastanozem) from those of neutral or alkaline soils, having lower SOC contents (Chernozem, Luvisol and Kastanozem). CAP2 axis was associated with  $T_{max}$  ( $r = -0.40$ ), BD ( $r = -0.34$ ), pH ( $r = -0.27$ ), ppt ( $r = 0.41$ ), actual evapotranspiration (aet,  $r = 0.39$ ) and runoff (q,  $r = 0.39$ ). This axis



**TABLE 3** CAP analyses based on Mahalanobis distance for MG-RAST Functional profiles annotations (level 2 Subsystem) and P-Enzyme relative abundance (A) and based on Kantorovich-Rubinstein distance for each of eight enzymes (B)

Variables	Subsystems		Relative abundance P-enzymes		Subsystems		Relative abundance P-enzymes		Variables	
	p value	CAP1 (r)	CAP1 (r)	Variables	p value	CAP1 (r)	CAP1 (r)	Variables	p value	Variables
pH	0.001	-0.74	-0.49	0.001	0.001	-0.70	-0.27			
SOC	0.001	0.56	-0.22	0.009	0.52	0.21	0.12			
Pav	0.059	0.57	-0.29	0.004	0.52	0.04	0.04			
CEC	0.057	0.15	-0.16	0.086	0.28	-0.48	0.18			
BD	0.001	-0.52	-0.12	0.001	0.12	-0.4	-0.02			
Clay	0.001	-0.04	0.23	0.003	0.12	0.49	-0.12			
Sand	0.001	0.12	0.07	0.001	-0.4	0.67	0.38			
Silt	0.070	-0.14	-0.31	0.001	0.49	0.72	0.39			
aet	0.001	0.65	0.18	0.030	0.67	0.82	0.12			
q	0.001	0.57	0.49	0.002	0.72					
Mositure	0.002	0.54	0.48	0.079	0.82					
<b>(B)</b>										
Variables										
Nsap—A			Nsap—B			Nsap—C			PhoA	
p value	CAP1 (r)	CAP2 (r)	p value	CAP1 (r)	CAP2 (r)	p value	CAP1 (r)	CAP2 (r)	p value	PhoX
0.028	0.06	0.27	0.120	0.47	-0.28	0.012	0.14	0.12	0.074	PhoD
0.045	-0.11	-0.18	0.523	-0.64	-0.19	0.016	0.02	-0.26	0.100	p value
0.197	-0.18	-0.17	0.186	-0.60	-0.27	0.106	-0.02	-0.28	0.325	CAP2 (r)
0.001	-0.38	0.23	0.907	-0.43	-0.07	0.158	0.02	-0.01	0.015	CAP1 (r)
0.026	0.09	0.13	0.002	0.65	0.16	0.034	0.07	0.43	0.433	% variance
0.061	0.00	-0.35	0.649	-0.30	0.140	0.001	0.18	-0.69	0.070	p value
0.002	0.17	0.39	0.027	0.40	-0.05	0.033	-0.02	0.53	0.571	CAP2 (r)
0.004	-0.34	-0.26	0.467	-0.34	-0.06	0.101	-0.16	-0.13	0.460	CAP1 (r)
0.001	0.01	-0.43	0.848	-0.59	0.32	0.001	0.17	-0.21	0.015	% variance
0.050	-0.01	-0.46	0.002	-0.80	0.27	0.023	0.00	-0.36	0.582	p value
0.071	-0.27	-0.51	0.158	-0.68	0.19	0.044	-0.17	-0.45	0.580	CAP2 (r)
0.082	0.00	-0.48	0.405	-0.75	0.32	0.009	0.09	-0.30	0.458	CAP1 (r)
0.001	-0.24	-0.17	0.508	0.09	0.31	0.001	-0.50	-0.04	0.057	% variance
CAP model	0.001	33.70	0.050	32.60	0.001	35.70	13.00	49.80	0.001	41.40
CAP 1 axis	0.001	11.54	0.001	11.60	0.001	13.90	8.70	19.80	0.001	19.00
CAP 2 axis	0.01	4.20	0.890	4.10	0.009	4.70	3.90	12.90	0.001	4.70

TABLE 3 (Continued)

Variables	Subsystems		Relative abundance P-enzymes		Subsystems		Relative abundance P-enzymes	
	p value		CAP1 (r)	Variables	p value		CAP1 (r)	Variables
Variables								
		BPP			Cphy			
	p value		CAP1 (r)		p value		CAP1 (r)	CAP2 (r)
pH	0.001		0.440	CAP2 (r)	0.010	0.570	0.290	0.290
SOC	0.066		-0.280		0.193	-0.260	-0.180	-0.180
Pav	0.048		-0.330		0.007	-0.340	-0.160	-0.160
CEC	0.007		-0.370		0.010	-0.320	-0.420	-0.420
BD	0.001		0.270		0.122	0.360	0.100	0.100
Clay	0.022		-0.090		0.526	0.190	-0.040	-0.040
Sand	0.001		0.280		0.222	-0.110	0.130	0.130
Silt	0.004		-0.350		0.302	0.030	-0.160	-0.160
aet	0.012		-0.310		0.928	-0.490	-0.470	-0.470
q	0.01		-0.370		0.002	-0.330	-0.380	-0.380
Mositure	0.169		-0.620		0.030	-0.680	-0.400	-0.400
ppt	0.164		-0.360		0.420	-0.440	-0.460	-0.460
T <sub>max</sub>	0.001		-0.290		0.508	0.150	0.000	0.000
	p value	% variance			p value	% variance		
CAP model	0.001	47.000			0.030	49.500		
CAP 1 axis	0.001	23.000			0.010	10.700		
CAP 2 axis	0.001	5.000			0.100	7.6000		

Note: Environmental variables included: pH, Soil Organic Carbon (SOC), Estimated organic available P (Pav), Cation Exchange Capacity (CEC), Bulk Density (BD), Clay, Sand, silt, actual evapotranspiration (aet), runoff (q), soil moisture (moisture), precipitation (ppt) and maximum Temperature (T<sub>max</sub>) (Table 3B). PERMANOVA analysis with 999 permutations was performed to determine the significance between the sites/MG-RAST project. The values marked with red are significant at  $p < 0.01$ .



separated soil metagenomes associated with lower aet values and relatively high  $T_{\max}$ . All variables were significant with a  $p < 0.001$  (Table 3A).

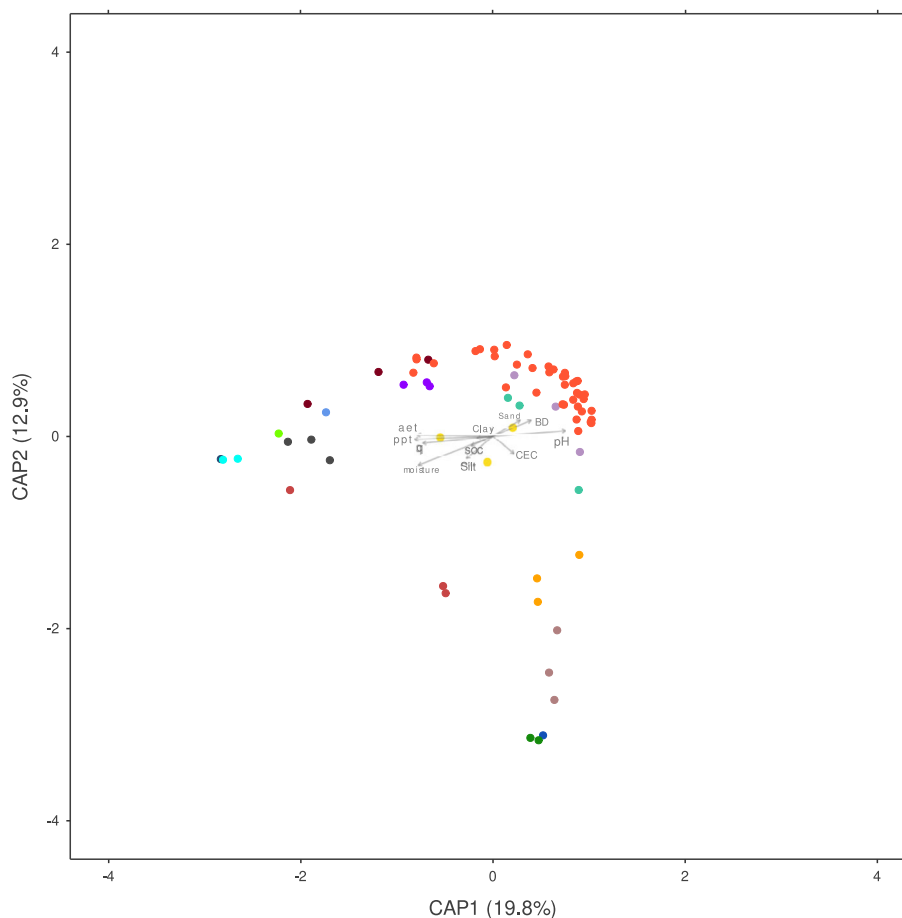
## Analyses on abundance and phylogeny of P-enzyme coding genes

To gain deeper insight into the diversity and abundance of P-enzymes coding genes, we performed CAP analyses using Kantorovich–Rubinstein (KR-CAP) distance matrices between samples to include not only abundance but also phylogenetic information.

The *phoD* KR-CAP analysis explained 49.8% of the total variance in the data set ( $p < 0.001$ ). Eleven out of 13 environmental variables were associated with the first two KR-CAP axes. The KR-CAP1 axis was negatively associated with pH ( $r = -0.75$ ), BD ( $r = -0.39$ ), Sand ( $r = -0.28$ ), CEC ( $r = -0.22$ ) and positively with

ppt ( $r = 0.80$ ), aet ( $r = 0.79$ ), q ( $r = 0.72$ ),  $T_{\max}$  ( $r = 0.313$ ), Silt ( $r = 0.27$ ) and SOC ( $r = 0.22$ ). This axis separated soils with low pH, relative high values of SOC and  $T_{\max}$  (Cambisols, Ferrasols, Mollisols/Phaeozem, Luvisol/Kastanozem) from soils with higher pH and lower  $T_{\max}$ . The KR-CAP2 axis was characterized by a negative association with  $T_{\max}$  ( $r = -0.41$ ), soil moisture ( $r = -0.29$ ) and Silt ( $r = -0.22$ ). This axis separated soil with neutral pH and relatively high Silt and Sand values from the rest of the samples (Table 3B and Figure 2).

We performed the same analysis for the rest of the P-enzymes coding genes and the results are summarized in Table 3B (Figures S5–S7). Notably, pH,  $T_{\max}$  and aet were associated with all P-enzyme coding gene distributions. SOC displayed a high correlation with alkaline phosphatases *phoD* and *phoX*, and acid phosphatases Nsap-A and Nsap-C coding genes, and estimated organic available P (Pav) was mainly



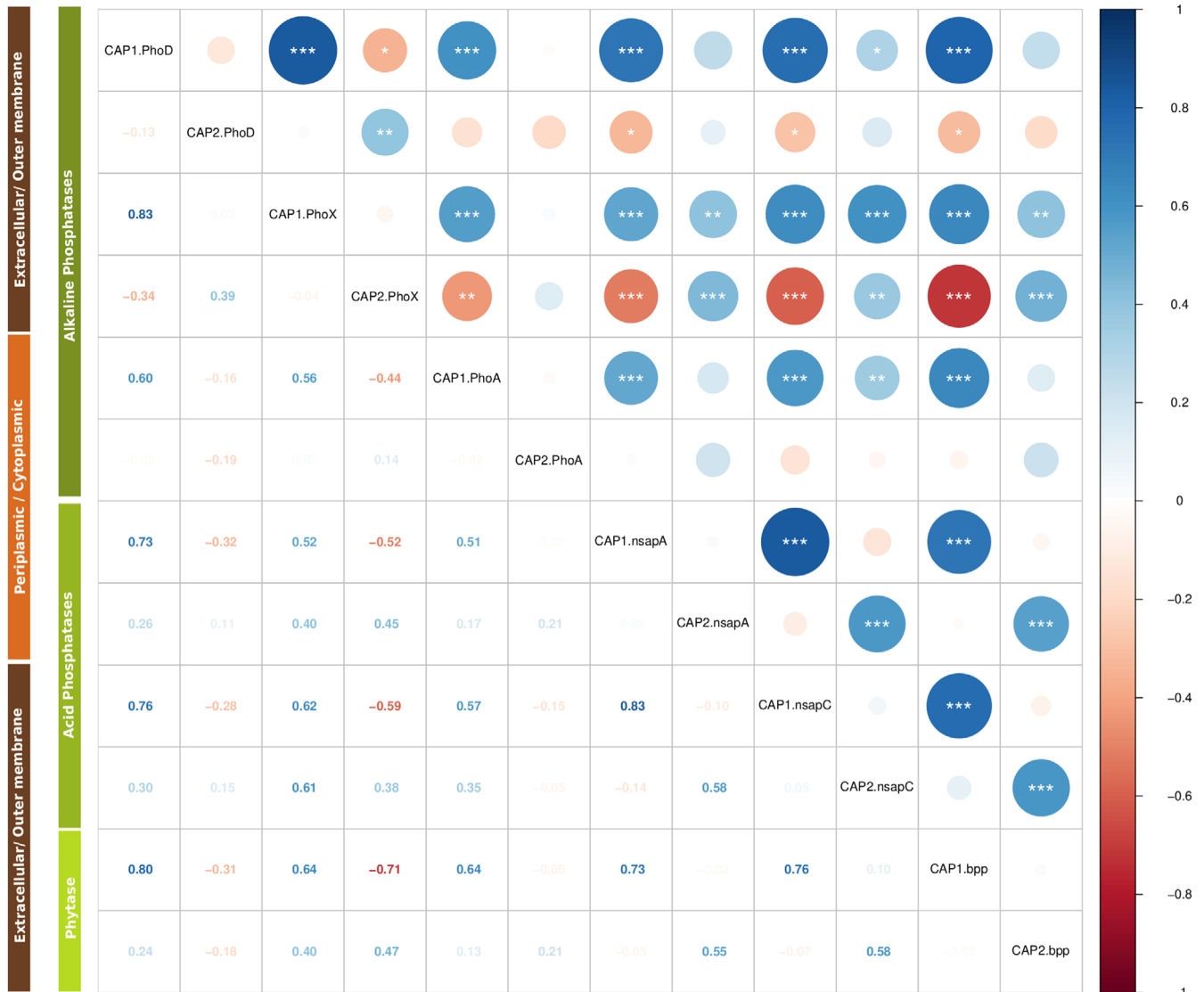
**FIGURE 2** CAP based on Kantorovich-Rubinstein distance for *phoD*. PERMANOVA analysis with 999 permutations was performed to determine the significance between the sites/MG-RAST project. For each MG-RAST project, three samples with the same geo-reference were included. Each point represents samples from the project mpg1992 (blue); mpg3520 (green); mpg5588 (dark red); mpg7792 (grey); mpg8624 (mustard); mpg9904 (violet); mpg10450 (dark blue); mpg10523 (stone blue); mpg10541 (turquoise); mpg10956 (yellow); mpg13011 (lilac); mpg13520 (jade); mpg13948 (orange); mpg20922 (brown); mpg89409 (brick-red); mpg91922 (light green); mpg93346 (light blue). Vector lengths represent the correlation between each variable and the axes. CAP analysis was performed with the Vegan R package and graphics were produced with the R package ggplot2.

associated with the phytases coding genes. Next, CEC showed a high correlation with alkaline phosphatases and phytase coding genes. Finally, clay content was related mainly to alkaline phosphatase coding genes (Table 3B).

### Covariation of P-enzymes genes

To examine the co-variation between P-enzymes we compared their corresponding KR-CAP analyses

results. We first limited the analysis to the P-enzymes genes present in at least 50 samples (all but Nsap-B and Cphy coding genes). All first KR-CAP axes showed a highly significant positive correlation between them (Figure 3 and Table S7). The second axes of this analysis showed a different behaviour, *phoA* KR-CAP2 showed no correlation with any other axis, *phoD* and *phoX* KR-CAP2 displayed a similar trend between them and with a rather idiosyncratic relationship with the rest of the genes. Finally, Nsap-A (*phoC*), Nsap-C (*olpA*) and BPP (*phyL* and *phyS*) coding genes displayed a



**FIGURE 3** Correlation matrix of KR-CAP axes. (A) Correlogram of the alkaline phosphatases displays the Pearson correlation coefficients between KR-CAP PhoD axes, KR-CAP PhoX, and KR-CAP PhoA; KR-CAP PhoX and KR-CAP PhoA. The correlation coefficients are coloured according to their values; blue is the positive values and red is the negative values. (B) Correlogram of the acid phosphatases displays the Pearson correlation coefficients between KR-CAP Nsap-A axes, KR-CAP Nsap-B and KR-CAP Nsap-C; KR-CAP Nsap-B and KR-CAP Nsap-C. The correlation coefficients are coloured according to their values; blue is the positive values and red is the negative values. (C) Correlogram of the phytases displays the Pearson correlation coefficients between KR-CAP BPP axes and KR-CAP CPhy. The correlation coefficients are coloured according to their values; blue is the positive values and red is the negative values. (D) Correlogram of the most abundance displays the Pearson correlation coefficients. The correlation coefficients are coloured according to their values; blue is the positive values and red is the negative values. Correlation analysis and graphics were performed with the cor R package.

very similar response (Figure 3 and Table S7). No specific trend was observed according to the predicted cellular localization of the proteins.

Nsap-B and Cphy coding genes were present in fewer samples, so we compared them to the other non-specific acid phosphatase and phytase, respectively. For Nsap-B, the KR-CAP1 axis was significantly correlated to the KR-CAP2 axes from the other two Nsap (Figure S8 and Table S7). In addition, the CPhy KR-CAP2 axis was correlated to the BPP KR-CAP1 axis (Figure S8 and Table S7).

Finally, when comparing each P-enzyme CAP analysis with the subsystem level CAP analysis we observed that only *phoD* (CAP1-PhoD vs. CAP2-SS = 0.44), BPP (*phyL* and *phyS*) (CAP1-BPP vs. CAP1-SS = -0.41, CAP2-BPP vs. CAP2-SS = 0.55) and Cphy (*phyA*) (CAP1-Cphy vs. CAP1-SS = -0.59), displayed significant correlations between the axes (Table S8).

## Edge-PCA and taxonomic identification of differentially observed P-enzymes coding genes

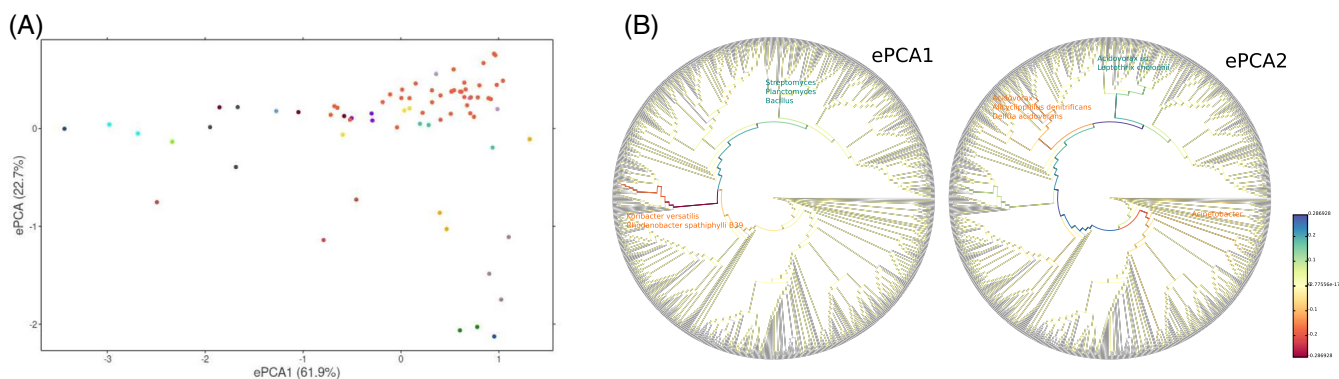
Edge-PCA analysis was applied to examine the variation in phylogenetic diversity of P-enzyme coding genes among the soil metagenomes; a summary of the results is shown in Table S9. It is important to note that the first and second edgePCA components were highly correlated with the first and second KR-CAP axes (except for the low abundance genes encoding *phoA* and Nsap-B *aphA*), this enables us to connect the environmental variables to specific lineages of each gene.

In the *phoD* analysis, the first edge-PCA axis separated samples by soil type, pH and SOC content. The

differences showed that the gene variants of the species *Koribacter versatilis* (class *Acidobacteriia*) and *Rhodanobacter spathiphylli* (class *Gammaproteobacteria*) (Figure 4B) were more abundant in soils classified as Ferrasols, Cambisols, Molisols/Phaeozem and Vertisol/Phaeozem with low pH and relatively high SOC content (left quadrant of Figure 4A). On the other hand, variants associated with *Actinomyces*, *Bacillus* and *Planctomyces* (Figure 4B) were more abundant in Kastanozem, Chernozem, Luvisol and Fluvisols soils with higher pH (ranged to 7.5) and lower SOC content (right quadrant of Figure 4A). The second axis was associated with *phoD* coding genes harboured by *Burkholderiales* and *Acinetobacter* with higher abundance in soils with neutral pH and low clay content (Tables S1 and S9 and Figure 4A).

The alkaline phosphatases *phoX* and *phoA* showed a narrower phylogenetic distribution and *Alphaproteobacteria* (*Rosevivax* and *Agrobacterium* among others) genes were predominant in soils with high SOC values and relatively high  $T_{max}$  (23°C) (Table S1 and Table S9). The *Burkholderiales* variants were observed in soil samples with near-neutral pH and average SOC and CEC values (Figure S9). The genes *phoA* of *Pantoea* and *Providencia* together with *Acinetobacter* and *Actinobacter* genera were associated with varying abundance between samples (Figure S9). Again, *Acinetobacter* was differential and more abundant in soils with circum-neutral pH and average SOC and CEC values (Figure S9).

We identified the acid phosphatases Nsap-A coding genes harboured by *Pedospaera*, *Dyella jiangningensis* and *Dyella japonica* as the differentials and the most abundant among soils with average SOC and CEC values and sandy texture (Figure S9). On the other hand, *Sphingomonas* sp., *Phenylobacterium* sp.,



**FIGURE 4** (A) Graphic representation of the first two axes of the edge-PCA for *phoD* using samples as observations. Each point represents samples from the project mpg1992 (blue); mpg3520 (green); mpg5588 (dark red); mpg7792 (grey); mpg8624 (mustard); mpg9904 (violet); mpg10450 (dark blue); mpg10523 (stone blue); mpg10541 (turquoise); mpg10956 (yellow); mpg13011 (lilac); mpg13520 (jade); mpg13948 (orange); mpg20922 (brown); mpg89409 (brick-red); mpg91922 (light green); mpg93346 (light blue). (B) The phylogeny distribution of *phoD* hits along the first and second axis of the analysis (proteins with positive coefficients are marked in blue and proteins with negative coefficients are marked in orange). The edge-PCA was performed using gappa software and tree and domain composition diagrams were drawn using Archaeopteryx (<https://sites.google.com/site/cmzmasek/home/software/forester>).

*Rhodanobacter* sp. and *Caulobacter* species variants were more abundant in soils with high clay content (Figure S9).

The Nsap-C coding genes harbouring by *Stenotrophomonas* (*Gammaproteobacteria*) and *Pedobacter* genera were differential and the most abundant in soils with low values of soil moisture and actual evapotranspiration (aet). *Enterobacter* Nsap-B coding genes variants predominated in soils classified as Ferralsols, Andosols and Luvisol with acidic pH (range to 5), and high aet and precipitation (ppt) values. Metagenomes from Fluvisol, with a pH = 7, were associated with P-enzymes gene encoding variants of *Photobacter* and *Marinomonas* and were strikingly different from the rest (Figure S9).

We only found the BPP phytase coding genes in soil samples with pH values above 6.6. BPP coding gene variants of the *Acinetobacter*, *Pseudomonas*, *Methylophaga*, *Pseudoalteromonas* and *Alteromonadales* (*Gammaproteobacteria*), and *Shewanella* and *Hylemonella* (*Betaproteobacteria*), dominate in clay soils with high CEC values. BPP genes harboured by *Bacillus* species were most abundant in sandy soils with low nutrient content.

CPhy coding genes from *Beta* and *Deltaproteobacteria*, *Clostridia* and several genera of *Negativicutes* classes varied across the samples but there was no clear signal to reveal associations with environmental variables.

## DISCUSSION

Soil ecosystems include complex interrelations among different factors including soil types, plant communities, microbial communities (bacteria, fungi, archaea, viruses and protozoa), macro and micro fauna, environmental variables, etc. (Islam et al., 2020). The present work focused on the Bacterial fraction of the soil microbial community from grassland biomes, in particular the abundance and phylogenetic diversity of P-enzyme coding genes from the grassland biomes, using a metagenomic approach. The analysed samples represent different environmental conditions defined by the physical and chemical soil properties, and climate variables (Amundson, 2013; Islam et al., 2020). We included publicly available data from MG-RAST and other sources for each project/sample. Other interesting data, such as the composition of the plant community or short-term/long-term experiments, was not included, which can constitute an interesting input for the analyses and discussion of our study.

Microbial P enzymes, such as phosphatases (Nannipieri et al., 2011; Rodriguez et al., 2006) and phytases (Tan et al., 2013; Yao et al., 2012), play a crucial role in the phosphorus cycle by participating in the release of Pi from organophosphorus compounds, the last step of the P cycle (Zeng et al., 2022). One

valuable result of this study is that it confirms and expands the idea of the large variability in abundance and diversity of P-enzymes coding genes within grassland ecosystems across the planet.

Our analyses showed that the alkaline phosphatases were the most abundant P-enzymes genes in the whole dataset, being the *phoD* gene the most abundant and also with the widest phylogenetic distribution, regardless of the soil properties. This result is in accordance with reports by previous, but more restricted, metagenomic studies where this gene was the most frequently alkaline phosphatase found in different soils (Bergkemper et al., 2016; Park et al., 2022; Tan et al., 2013). The *phoA* was less abundant, and the difference with *phoD* or *phoX* can result from the differences in substrate specificity and co-factor requirements between them. Bacterial cells may possess either *phoX* or *phoA* or both. They are presumed to have similar roles in facilitating access to a diverse array of phosphoester compounds and are more active against organic phosphates and nucleotides. Nonetheless, they may function at varying levels of substrate concentrations (Sebastian & Ammerman, 2011). A study about PhoA activity in marine ecosystems showed that this enzyme has an activity for mono-, di- and triesterase activity (Srivastava et al., 2021). PhoX is essential for utilizing monophosphate esters at low substrate concentrations in *Rhizobium pomeroyi* (Sebastian & Ammerman, 2011). The substrate specificity of PhoD is unknown. Some work has reported phosphodiesterase activity against cell wall teichoic acids and phospholipids (Bergkemper et al., 2016; Rodriguez et al., 2014). However, the contribution to the Pho-regulated phosphatase activity of *Pseudomonas fluorescens* does not seem to be significant (Monds et al., 2006). A new alkaline phosphatase, PafA, has recently been described in plant-associated Bacteroidetes (Lidbury et al., 2021). Unlike PhoD, PhoX and PhoA, this enzyme exhibits constitutive phosphatase activity and is fully functional in the presence of high phosphate concentrations with high monophosphatase activity. PafA plays a critical role in global biogeochemical cycles and has potential applications in sustainable agriculture (Lidbury et al., 2021).

On the other extreme, genes encoding Nspa-B and Cphy were scarce in the whole dataset. These genes tend to show weaker associations with the environmental variables and other P-enzyme coding genes. This could be due to the low numbers in which these genes appear or to genuine biological reasons.

## Environmental variables and P-enzyme coding genes abundance and diversity

We showed that several environmental variables are related to the diversity and abundance of P-enzyme coding genes.  $T_{max}$ , pH, SOC and soil moisture are



associated with alkaline phosphatase gene abundance. The *phoD* and *phoX* genes showed a high correlation with SOC and clay. Several recent studies report the effect of SOC, N and organic P content on the abundance and diversity of both enzymes and the corresponding bacteria (Li et al., 2021; Ragot et al., 2017; Wei et al., 2021). A local-scale study of three land uses with differential SOC (fallow, arable, grassland) demonstrated there was a positive correlation between alkaline phosphatases gene abundance and soil organic matter contents (Neal et al., 2017). In addition, the predicted extracellular location of both enzymes (Neal et al., 2017) may explain the importance of clay content in relation to its stabilization role, immobilization and maintenance of the enzymatic activity (Margalef et al., 2017). The *phoD* genes are widely distributed among different classes of Bacteria, in this study we found that variants associated with *Koribacter* (*Acidobacteria* class) and *Rhodanobacter* genus (*Gammaproteobacteria* class) were more abundant in soils with relatively high SOC values and low pH. These variants have been identified as a dominant phylotype in arable silty clay loam soil Chromic Luvisol in the United Kingdom (Neal et al., 2017). The second one also has been identified as a dominant phylotype in the rhizospheres of maize and sorghum in a Brazilian Distroferic Red Latosol (Neal et al., 2021). Both bacterial species represent classes that possess a comprehensive set of genes that allow them to use a wide variety of substrates, responding efficiently to environmental changes and conferring their ability to adapt to various ecological niches (Kalam et al., 2020; Kurm et al., 2017). Variants associated with *Bacillus*, *Actinomyces* and *Planctomyces* were prevalent in soils with lower SOC and neutral pH. The last two species have been found dominant in soils with low nutrient content, even the *Planctomyces* showed a negative correlation with this variable (Garaycochea et al., 2020; Hermans et al., 2017; Lewin et al., 2017). Nevertheless, the main driver that explains the difference in species abundances appears to be pH, since all reported species are heterotrophs (Kielak et al., 2016; Saxena et al., 2020). The *phoX* gene represented by the *Burkholderia* genus was preferred in soils with low and medium content of SOC and neutral pH. Bacteria from this genus present a wide repertoire of metabolic pathways making them more competitive in nutrient-restrictive environments, since they are capable to degrade recalcitrant compounds, and unlike most Bacteria, *Burkholderia* species are more competitive in low and moderate pH conditions (Morya et al., 2020; Stopnisek et al., 2014).

Regarding acid phosphatases, the Nsap-A coding genes were found in *Dyella* and *Rhodanobacter* genera. These species use different carbon sources and have been reported to be dominant in acid and neutral soils (Dahal & Kim, 2017; Weon et al., 2009). On the

other hand, the Nsap-C coding gene was identified in *Alpha* and *Gammaproteobacteria*, *Flavobacteria* and *Sphingobacteria* classes, consistent with previous evidence (Gaiero et al., 2020; Neal et al., 2017). The proportion of both non-specific acid phosphatases found in the grassland set studied here was similar to that reported for UK grassland soils (Neal et al., 2017). The predominance of acid phosphatases in grassland could be influenced by the interaction between microorganisms and plant communities, as both are capable of producing these enzymes (Mhlongo et al., 2018). The observed proportion of Nsap-B is similar to that reported by Udaondo et al. (2020), who not only found that this enzyme was less abundant in different niches but also that it was restricted to a limited number of microbial families, some of which were pathogens.

In the cases of phytases, BPP coding genes showed an abundance and phylogenetic distribution in accordance with what has been reported. The BPP coding genes are widespread and are distributed among various species of soil bacteria (Huang et al., 2009; Jorquera et al., 2008; Kumar et al., 2017; Lim et al., 2007). However, some studies have observed that the presence of BPP coding genes is rare in *Betaproteobacteria* (Cotta et al., 2016), we found that the BPP coding genes variants were mainly from *Bacillales* and *Beta* and *Gammaproteobacteria*. The BPP coding genes in this study were found restricted to soil with pH above 6.6, which is in accordance with what was reported, particularly in several strains from the *Bacillus* genus, where the BPPs enzymes are optimally active at pH 6.0–7.5 (Cheng & Lim, 2006; Farhat et al., 2008; Huang et al., 2009; Kerovuo et al., 1998; Kumar et al., 2017). On the other hand, the Cphy coding gene was the least abundant enzyme in the grasslands metagenomes, contrary to those found by Neal et al., 2017 where Cphy tended to have a similar abundance that BPP in the studied grasslands from the UK.

The pH appears as an important factor associated with both acid and alkaline phosphatases, as well as phytases, abundance and diversity. Even though, our results show a global trend of an increase in the genes encoding these enzymes (PhoD, PhoX, Nspa-C and BPP) with pH, all enzymes are relatively abundant in the pH range covered in this study, rendering it difficult to test a direct association between the enzyme classification (as acid or alkaline) and the soil pH.

It is important to bear in mind that the taxonomic associations of each gene sequence are dependent on the database, is clear that including different sequences of more taxa might result in the discovery of new variants and/or better assignments of the sequences. Nevertheless, many of the results here obtained will still hold being enriched with the new putative ones.

## Co-variation of P-enzyme coding genes

We have shown a strong relationship in the abundance and diversity patterns between the different P-enzymes herein studied. Indeed, KR-CAP analyses show strong correlations between them, somehow weaker in the less abundant genes. These results uncover a somehow intuitive result. We should bear in mind that we are counting the aggregate of each gene in a whole community, thus, variation in abundance and diversity of a given gene is the product of a change at the community level. So the process of selection in the assemblage of each community is a balance between how each organism crafted its genome and the interaction between them and the environment. The high correlation between KR-CAP analyses, which involve abundance, diversity and environmental variables, suggests a tight relationship between the P-enzyme genes. This implies that for each environmental condition, the way each P-enzyme gene contributes to phosphorus cycling and metabolism is connected to the rest of them (in both abundance and diversity). The different taxa that appear associated with each P-enzyme coding gene in the edgePCA analysis (Table S9) are indicative that different organisms are contributing to the P-enzyme gene pool.

Another interesting result was the association of P-enzymes with the general functional profiles (Table S8). Here, the results are somehow at odds with the previous one. The first axes of PhoD, PhoX, PhoA and BPP coding genes were strongly correlated with the CAP2 of the functional profiles. Nevertheless, Cphy and Nsap-X genes, showed no correlation, suggesting that there could be some variability in this respect.

One important question is to understand if the P-enzymes are driven particularly by the change of certain organisms that are carrying them or, in turn, they are following the general major changes in the community structure. One possible hint in this direction is given by the previous comparison, indicating that these P-enzymes genes may be accompanying the general change in the functional structure of the metagenome, whilst there is room for a more idiosyncratic manner. Nevertheless, more studies should be carried out to gain deeper insight into this interesting and complex question.

## Concluding remarks

The environmental variables explained a relatively low proportion of the variability in bacterial functional profiles. The use of information from samples from very distant sites determines only the effect on the diversity of the variables with greater differences among the sites. However,  $T_{max}$ , soil pH and evapotranspiration were related to the abundance and diversity of almost

the eight key enzymes involved in P organic cycling. Likewise, it was possible to identify the effect of other variables with a more localized effect, such as soil texture and soil organic content, as important determinants of microbial community structure and functions. The complexity of the studied system requires a combination of approaches and the generation of local data that allow the understanding of factors affecting the presence of bacteria carrying P-enzymes genes as well as their functionality and to integrate these results into a broader scale to detect global patterns of diversity that could potentially lead to better understanding and management of soil P cycling.

## AUTHOR CONTRIBUTIONS

**Silvia Garaycochea:** Conceptualization (equal); data curation (equal); formal analysis (equal); funding acquisition (supporting); investigation (equal); methodology (equal); resources (equal); writing – original draft (equal); writing – review and editing (equal). **Nora Adriana Altier:** Conceptualization (supporting); funding acquisition (lead); project administration (lead); writing – original draft (supporting). **Carolina Leoni:** Formal analysis (supporting); methodology (supporting); writing – original draft (supporting). **Andrew L. Neal:** Conceptualization (supporting); methodology (supporting); writing – original draft (supporting). **Héctor Romero:** Conceptualization (equal); data curation (equal); formal analysis (equal), methodology (supporting), writing – original draft (equal), writing – review and editing (equal).

## ACKNOWLEDGEMENTS

We acknowledge Victoria Bonnacarrère for their contribution to the discussion. This work was funded by the Instituto Nacional de Investigación Agropecuaria, INIA (Project SA47, SA 26 and SA 24), Agencia Nacional de Investigación e Innovación (POS NAC 2015 1 110075).

## CONFLICT OF INTEREST

The authors declare no conflict of interest.


## DATA AVAILABILITY STATEMENT

All data used in this work are publicly available. Additional data, additional results and scripts are available at <https://github.com/eletor-uy/Grasslands>.

## ORCID

*Silvia Garaycochea*  <https://orcid.org/0000-0003-1629-0746>

*Nora Adriana Altier*  <https://orcid.org/0000-0002-3173-6016>

*Carolina Leoni*  <https://orcid.org/0000-0002-3891-564X>

*Andrew L. Neal*  <https://orcid.org/0000-0003-4225-1396>

## REFERENCES

- Achat, D.L., Pousse, N., Nicolas, M., Brédoire, F. & Augusto, L. (2016) Soil properties controlling inorganic phosphorus availability: general results from a national forest network and a global compilation of the literature. *Biogeochemistry*, 127, 255–272. Available from: <https://doi.org/10.1007/s10533-015-0178-0>
- Amadou, I., Houben, D. & Faucon, M.-P. (2021) Unravelling the role of rhizosphere microbiome and root traits in organic phosphorus mobilization for sustainable phosphorus fertilization. *A review Agronomy*, 11, 2267. Available from: <https://doi.org/10.3390/agronomy11112267>
- Amundson, R. (2013) Soil formation. In: Holland, H.D. & Turekian, K. K. (Eds.) *Treatise on geochemistry*, 2nd edition. USA: Elsevier Science, pp. 1–26. Available from: <https://doi.org/10.1016/B978-0-08-095975-7.00501-5>
- Anderson, M.J. & Willis, T.J. (2003) Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. *Ecology*, 84(2), 511–525. Available from: [https://doi.org/10.1890/0012-9658\(2003\)084\[0511:CAOPCA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2)
- Ariza, A., Moroz, O.V., Blagova, E.V., Turkenburg, J.P., Waterman, J., Roberts, S.M. et al. (2013) Degradation of Phytate by the 6-Phytase from *Hafnia alvei*: A combined structural and solution study. *PLoS One*, 8(5), e65062. Available from: <https://doi.org/10.1371/journal.pone.0065062>
- Awasthi, R., Tewari, R. & Nayyar, H. (2011) Synergy between plants and P solubilizing microbes in soils: effects on growth and physiology of crops. *International Research Journal of Microbiology*, 2, 484–503.
- Azeem, M., Riaz, A., Nawaz, A., Hayat, R., Hussain, Q., Tahir, N. et al. (2015) Microbial phytase activity and their role in organic P mineralization. *Archives of Agronomy and Soil Science*, 61(6), 751–766. Available from: <https://doi.org/10.1080/03650340.2014.963796>
- Blair, J., Nippert, J. & Briggs, J. (2014) Grassland Ecology. *Ecology and the Environment*, 389–423. Available from: [https://doi.org/10.1007/978-1-4614-7501-9\\_14](https://doi.org/10.1007/978-1-4614-7501-9_14)
- Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T. et al. (2019) EPA-ng: massively parallel evolutionary placement of genetic sequences. *Systematic Biology*, 68, 365–369. Available from: <https://doi.org/10.1093/sysbio/syy054>
- Barnett, K.L. & Facey, S.L. (2016) Grasslands, invertebrates, and precipitation: A review of the effects of climate change. *Frontiers in Plant Science*, 7, 1196. Available from: <https://doi.org/10.3389/fpls.2016.01196>
- Bergkemper, F., Kublik, S., Lang, F., Krüger, J., Vestergaard, G., Schloter, M. et al. (2016) Novel oligonucleotide primers reveal a high diversity of microbes which drive phosphorus turnover in soil. *Journal of Microbiological Methods*, 125, 91–97. Available from: <https://doi.org/10.1016/j.mimet.2016.04.011>
- Braithwaite, D.T. & Keegan, K.P. (2018) matR: metagenomics analysis tools. R Package Version 0.9.1. <https://CRAN.R-project.org/package=matR>
- Caffaro, M., Balestrasse, K. & Rubio, G. (2020) Adsorption to soils and biochemical characterization of commercial phytases. *The Soil*, 6, 153–162. Available from: <https://doi.org/10.5194/soil-6-153-2020>
- Cai, S., Deng, K., Tang, J., Sun, R., Lu, H., Li, J. et al. (2021) Characterization of extracellular phosphatase activities in periphytic biofilm from paddy field. *Pedosphere*, 31, 116–124. Available from: [https://doi.org/10.1016/S1002-0160\(20\)60061-3](https://doi.org/10.1016/S1002-0160(20)60061-3)
- Cheng, C. & Lim, B.L. (2006) Beta-propeller phytases in the aquatic environment. *Archives of Microbiology*, 185(1), 1–13. Available from: <https://doi.org/10.1007/s00203-005-0080-6>
- Condon, L.M., Turner, B.L. & Cade-Menun, B.J. (2005) Chemistry and dynamics of soil organic phosphorus. *Phosphorus: Agriculture and the Environment*, 46, 87–121. Available from: <https://doi.org/10.2134/agronmonogr46.c4>
- Cotta, S.R., Dias, A.C.F., Seldin, L., Andreote, F.D. & van Elsas, J.D. (2016) The diversity and abundance of phytase genes ( $\beta$ -propeller phytases) in bacterial communities of the maize rhizosphere. *Letters in Applied Microbiology*, 62(3), 264–268. Available from: <https://doi.org/10.1111/lam.12535>
- Czech, L., Barbera, P. & Stamatakis, A. (2020) Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics*, 36, 3263–3265. Available from: <https://doi.org/10.1093/bioinformatics/btaa070>
- Dahal, R.H. & Kim, J. (2017) *Rhodanobacter humi* sp. nov., an acid-tolerant and alkali-tolerant gammaproteobacterium isolated from forest soil. *International Journal of Systematic and Evolutionary Microbiology*, 67, 1185–1190. Available from: <https://doi.org/10.1099/ijsem.0.001786>
- de Lacerda, J.R.M., da Silva, T.F., Vollú, R.E., Marques, J.M. & Seldin, L. (2016) Generally recognized as safe (GRAS) *Lactococcus lactis* strains associated with *Lippia sidoides* Cham. are able to solubilize/mineralize phosphate. *Springerplus*, 5(1), 828. Available from: <https://doi.org/10.1186/s40064-016-2596-4>
- Dubeux, J., Sollenberger, L.E., Mathews, B.W., Scholberg, J.M. & Santos, H.Q. (2007) Nutrient cycling in warm-climate grasslands. *Crop Science*, 47, 915–928. Available from: <https://doi.org/10.2135/cropsci2006.09.0581>
- Evans, S.N. & Matsen, F.A. (2012) The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 74, 569–592. Available from: <https://doi.org/10.1111/j.1467-9868.2011.01018.x>
- Farhat, A., Chouayekh, H., Ben Farhat, M., Bouchaala, K. & Bejar, S. (2008) Gene cloning and characterization of a thermostable phytase from *Bacillus subtilis* US417 and assessment of its potential as a feed additive in comparison with a commercial enzyme. *Molecular Biotechnology*, 40(2), 127–135. Available from: <https://doi.org/10.1007/s12033-008-9068-1>
- Fraser, T.D., Lynch, D.H., Gaiero, J., Khosla, K. & Dunfield, K.E. (2017) Quantification of bacterial non-specific acid (pho C) and alkaline (pho D) phosphatase genes in bulk and rhizosphere soil from organically managed soybean fields. *Applied Soil Ecology*, 111, 48–56. Available from: <https://doi.org/10.1016/j.apsoil.2016.11.013>
- Gaiero, J., Bent, E., Boitt, G., Condon, L. & Dunfield, K. (2020) Effect of long-term plant biomass management on phosphatase-producing bacterial populations in soils under temperate grassland. *Applied Soil Ecology*, 151, 103583. Available from: <https://doi.org/10.1016/j.apsoil.2020.103583>
- Gaiero, J.R., Bent, E., Fraser, T.D., Condon, L.M. & Dunfield, K.E. (2018) Validating novel oligonucleotide primers targeting three classes of bacterial non-specific acid phosphatase genes in grassland soils. *Plant and Soil*, 427, 39–51. Available from: <https://doi.org/10.1007/s11104-017-3338-2>
- Garaycochea, S., Romero, H., Beyhaut, E., Neal, A.L. & Altier, N. (2020) Soil structure, nutrient status and water holding capacity shape Uruguayan grassland prokaryotic communities. *FEMS Microbiology Ecology*, 96, 12. Available from: <https://doi.org/10.1093/femsec/fiaa207>
- George, T.S., Giles, C., Menezes-Blackburn, D., Condon, L., Gama-Rodrigues, A.C., Jaisi, D. et al. (2018) Organic phosphorus in the terrestrial environment: A perspective on the state of the art and future priorities. *Plant and Soil*, 427, 191–208. Available from: <https://doi.org/10.1007/s11104-017-3391-x>
- Gerke, J. (2015) The acquisition of phosphate by higher plants: effect of carboxylate release by the roots. A critical review. *Journal of Plant Nutrition and Soil Science*, 178, 351–364. Available from: <https://doi.org/10.1002/jpln.201400590>
- Graça, J., Daly, K., Bondi, G., Ikoyi, I., Crispie, F., Cabrera-Rubio, R. & Schmalenberger, A. (2021) Drainage class and soil phosphorus availability shape microbial communities in Irish grasslands.



- European Journal of Soil Biology*, 104, 103297. Available from: <https://doi.org/10.1016/j.ejsobi.2021.103297>
- Gyaneshwar, P., Kumar, G.N., Parekh, L.J. & Poole, P.S. (2002) Role of soil microorganisms in improving P nutrition of plants. *Plant and Soil*, 245, 83–93. Available from: <https://doi.org/10.1023/A:1020663916259>
- Haygarth, P.M., Bardgett, R.D. & Condon, L.M. (2013) Nitrogen and phosphorus cycles and their management. In: *Soil conditions and plant growth*. Oxford: Blackwell Publishing Ltd, pp. 132–159. Available from: <https://doi.org/10.1002/9781118337295.ch5>
- Hermans, S.M., Buckley, H.L., Case, B.S., Curran-Cournane, F., Taylor, M. & Lear, G. (2017) Bacteria as emerging indicators of soil condition. *Applied and Environmental Microbiology*, 83(1), e02826-16. Available from: <https://doi.org/10.1128/AEM.02826-16>
- Huang, H., Shi, P., Wang, Y., Luo, H., Shao, N., Wang, G. et al. (2009) Diversity of beta-propeller phytase genes in the intestinal contents of grass carp provides insight into the release of major phosphorus from phytate in nature. *Applied and Environmental Microbiology*, 75, 1508–1516. Available from: <https://doi.org/10.1128/AEM.02188-08>
- Islam, W., Noman, A., Naveed, H., Huang, Z. & Chen, H. (2020) Role of environmental factors in shaping the soil microbiome. *Environmental Science and Pollution Research*, 27, 41225–41247. Available from: <https://doi.org/10.1007/s11356-020-10471-2>
- Jorquera, M., Martínez, O., Maruyama, F., Marschner, P. & de la Luz Mora, M. (2008) Current and future biotechnological applications of bacterial phytases and phytase producing bacteria. *Microbes and Environments*, 23, 182–191. Available from: <https://doi.org/10.1264/jisme.2.23.182>
- Kalam, S., Basu, A., Ahmad, I., Sayyed, R.Z., El-Enshasy, H.A., Dailin, D.J. et al. (2020) Recent understanding of soil acidobacteria and their ecological significance: A critical review. *Frontiers in Microbiology*, 11, 580024. Available from: <https://doi.org/10.3389/fmicb.2020.580024>
- Kahle, D. & Wickham, H. (2013) ggmap: Spatial Visualization with ggplot 2. *The R Journal*, 5(1), 144–161.
- Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 15, 3059–3066. Available from: <https://doi.org/10.1093/nar/gkf436>
- Kerovuo, J., Lauraeus, M., Nurminen, P., Kalkkinen, N. & Apajalahti, J. (1998) Isolation, characterization, molecular gene cloning, and sequencing of a novel phytase from *Bacillus subtilis*. *Applied and Environmental Microbiology*, 64(6), 2079–2085. Available from: <https://doi.org/10.1128/AEM.64.6.2079-2085.1998>
- Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., Van Veen, J.A. & Kuramae, E.E. (2016) The ecology of acidobacteria: moving beyond genes and genomes. *Frontiers in Microbiology*, 7, 744. Available from: <https://doi.org/10.3389/fmicb.2016.00744>
- Knapp, A.K., Harper, C.W., Danner, B.T. & Lett, M.S. (2002) Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science*, 298, 2202–2205. Available from: <https://doi.org/10.1126/science.1076347>
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood of phylogenetic inference. *Bioinformatics*, 35, 4453–4455. Available from: <https://doi.org/10.1093/bioinformatics/btz305>
- Kumar, V., Yadav, A.N., Verma, P., Sangwan, P., Saxena, A., Kumar, K. et al. (2017)  $\beta$ -Propeller phytases: diversity, catalytic attributes, current developments and potential biotechnological applications. *International Journal of Biological Macromolecules*, 98, 595–609. Available from: <https://doi.org/10.1016/j.ijbiomac.2017.01.134>
- Kurm, V., van der Putten, W.H., de Boer, W., Naus-Wiezer, S. & Hol, W.H.G. (2017) Low abundant soil bacteria can be metabolically versatile and fast growing. *Ecology*, 98, 555–564. Available from: <https://doi.org/10.1002/ecy.1670>
- Le Roux, X., Recous, S. & Attard, E. (2011) Soil microbial diversity in grasslands, and its importance for grassland functioning and services. In: Lemaire, G. & Chabbi, A. (Eds.) *Grassland productivity and ecosystem services*. Wallingford, UK: CABI International, pp. 158–165.
- Lenhart, P.A., Eubanks, M.D. & Behmer, S.T. (2015) Water stress in grasslands: dynamic responses of plants and insect herbivores. *Oikos*, 124, 381–390. Available from: <https://doi.org/10.1111/oik.01370>
- Letunic, I. & Bork, P. (2007) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23, 127–128. Available from: <https://doi.org/10.1093/bioinformatics/btl529>
- Lewin, G.R., Carlos, C., Chevrette, M.G., Horn, H.A., McDonald, B.R., Stankey, R.J. et al. (2017) Evolution and ecology of Actinobacteria and their bioenergy applications. *Annual Review of Microbiology*, 8(70), 235–254. Available from: <https://doi.org/10.1146/annurev-micro-102215-095748>
- Li, J., Xie, T., Zhu, H., Zhou, J., Li, C., Xiong, W. et al. (2021) Alkaline phosphatase activity mediates soil organic phosphorus mineralization in a subalpine forest ecosystem. *Geoderma*, 404, 115376. Available from: <https://doi.org/10.1016/j.geoderma.2021.115376>
- Lidbury, I.D.E.A., Borsetto, C., Murphy, A.R.J., Bottrill, A., Jones, A., Bending, G. et al. (2021) Niche-adaptation in plant-associated *Bacteroidetes* favours specialisation in organic phosphorus mineralisation. *The ISME Journal*, 15, 1040–1055. Available from: <https://doi.org/10.1038/s41396-020-00829-2>
- Lidbury, I.D.E.A., Fraser, T., Murphy, A.R.J., Scanlan, D.J., Bending, G.D., Jones, A.M.E. et al. (2017) The “known” genetic potential for microbial communities to degrade organic phosphorus is reduced in low-pH soils. *Microbiologypopen*, 6(4), e00474. Available from: <https://doi.org/10.1002/mbo3.474>
- Lim, B.L., Yeung, P., Cheng, C. & Hill, J.E. (2007) Distribution and diversity of phytate-mineralizing bacteria. *The ISME Journal*, 1, 321–330. Available from: <https://doi.org/10.1038/ismej.2007.40>
- Ma, B., Stirling, E., Liu, Y., Zhao, K., Zhou, J., Singh, B.K. et al. (2021) Soil biogeochemical cycle couplings inferred from a function-taxon network. *Research (Wash DC)*, 10(2021), 7102769. Available from: <https://doi.org/10.34133/2021/7102769>
- Margalef, O., Sardans, J., Fernández-Martínez, M., Molowny-Horas, R., Janssens, I.A., Ciais, P. et al. (2017) Global patterns of phosphatase activity in natural soils. *Scientific Reports*, 7, 1337. Available from: <https://doi.org/10.1038/s41598-017-01418-8>
- Matsen, F.A. & Evans, S.N. (2013) Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLoS One*, 8, e56859. Available from: <https://doi.org/10.1371/journal.pone.0056859>
- Meyer, F., Paarmann, D., D’Souza, M., Olson, R., Glass, E., Kubal, M. et al. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9(1). Available from: <https://doi.org/10.1186/1471-2105-9-386>
- Mhlongo, M.I., Piater, L.A., Madala, N.E., Labuschagne, N. & Dubery, I.A. (2018) The chemistry of plant-microbe interactions in the rhizosphere and the potential for metabolomics to reveal signaling related to defense priming and induced systemic resistance. *Frontiers in Plant Science*, 9, 112. Available from: <https://doi.org/10.3389/fpls.2018.00112>
- Minh, B., Schmidt, H., Chernomor, O., Schrempf, D., Woodhams, M., Von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530–1534. Available from: <https://doi.org/10.1093/molbev/msaa015>





- Misset, O. (2002) Phytase. In: Whitaker, J., Voragen, A. & Wong, D. (Eds.) *Handbook of Food Enzymology*. Boca Raton: CRC Press, pp. 687–706. Available from: <https://doi.org/10.1201/9780203910450>
- Monds, R.D., Newell, P.D., Schwartzman, J.A. & O'Toole, G.A. (2006) Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. *Applied and Environmental Microbiology*, 72(3), 1910–1924. Available from: <https://doi.org/10.1128/AEM.72.3.1910-1924.2006>
- Morrison, E., Newman, S., Bae, H.S., He, Z., Zhou, J., Reddy, K.R. et al. (2016) Microbial genetic and enzymatic responses to an anthropogenic phosphorus gradient within a subtropical peatland. *Geoderma*, 268, 119–127. Available from: <https://doi.org/10.1016/j.geoderma.2016.01.008>
- Morya, R., Salvachúa, D. & Thakur, I.S. (2020) Burkholderia: an untapped but promising bacterial genus for the conversion of aromatic compounds. *Trends in Biotechnology*, 38(9), 963–975. Available from: <https://doi.org/10.1016/j.tibtech.2020.02.008>
- Nannipieri, P., Giagnoni, L., Landi, L. & Renella, G. (2011) Role of phosphatase enzymes in soil. In: *Phosphorus in action*. *Soil Biology*. Heidelberg: Springer, pp. 215–243. Available from: [https://doi.org/10.1007/978-3-642-15271-9\\_9](https://doi.org/10.1007/978-3-642-15271-9_9)
- Neal, A.L., Blackwell, M.S.A., Akkari, E., Clark, I.M., Hirsch, P.R. & Guyomar, C. (2018) Phylogenetic distribution, biogeography and the effects of land management upon bacterial non-specific acid phosphatase gene diversity and abundance. *Plant and Soil*, 427, 175–189. Available from: <https://doi.org/10.1007/s11104-017-3301-2>
- Neal, A.L., McLaren, T., Lourenço Campolino, M., David, H., Marcos Coelho, A., de Paula, G. et al. (2021) Crop type exerts greater influence upon rhizosphere phosphohydrolase gene abundance and phylogenetic diversity than phosphorus fertilization. *FEMS Microbiology Ecology*, 97, 4. Available from: <https://doi.org/10.1093/femsec/fiab033>
- Neal, A.L., Rossmann, M., Brearley, C., Akkari, E., Guyomar, C., Clark, I. et al. (2017) Land-use influences phosphatase gene microdiversity in soils. *Environmental Microbiology*, 19, 2740–2753. Available from: <https://doi.org/10.1111/1462-2920.13778>
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. et al. (2019) Vegan: community ecology package. R package version 2.62. <https://CRAN.R-project.org/package=vegan>
- Oliverio, A.M., Bissett, A., McGuire, K., Saltonstall, K., Turner, B.L. & Fierer, N. (2020) The role of phosphorus limitation in shaping soil bacterial communities and their metabolic capabilities. *MBio*, 11(5), e01718-20. Available from: <https://doi.org/10.1128/mBio.01718-20>
- Park, Y., Solhtalab, M., Thongsomboon, W. & Aristilde, L. (2022) Strategies of organic phosphorus recycling by soil bacteria: acquisition, metabolism, and regulation. *Environmental Microbiology Reports*, 14(1), 3–24. Available from: <https://doi.org/10.1111/1758-2229.13040>
- Quiquampoix, H. & Mousain, D. (2005) Enzymatic hydrolysis of organic phosphorus. In: Turner, B.L., Frossard, E. & Baldwin, D. (Eds.) *Organic phosphorus in the environment*. Wallingford (UK): CABI, pp. 89–112. Available from: <https://doi.org/10.1079/9780851998220.0089>
- Ragot, S., Kertesz, M. & Bünemann, E. (2015) phoD alkaline phosphatase gene diversity in soil. *Applied and Environmental Microbiology*, 81, 7281–7289. Available from: <https://doi.org/10.1128/AEM.01823-15>
- Ragot, S., Kertesz, M., Mészáros, E., Frossard, E. & Bünemann, E. (2017) Soil *phoD* and *phoX* alkaline phosphatase gene diversity responds to multiple environmental factors. *FEMS Microbiology Ecology*, 93, 1. Available from: <https://doi.org/10.1093/femsec/fiw212>
- Rao, M.A., Violante, A. & Gianfreda, L. (1994) Catalytic behavior of acid phosphatase immobilized on clay minerals and organo-mineral complexes transactions. In: 15th world congress of soil science; volumen 3b. International soil science society, Mexico. Acapulco (Mexico): Mexican Society of Soil Science, 117–118.
- Richardson, A. & Simpson, R. (2011) Soil microorganisms mediating phosphorus availability. *Plant Physiology*, 156, 989–996. Available from: <https://doi.org/10.1104/pp.111.175448>
- Robinson, M., McCarthy, D. & Smyth, G. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. Available from: <https://doi.org/10.1093/bioinformatics/btp616>
- Rodriguez, F., Lillington, J., Johnson, S., Timmel, C.R., Lea, S.M. & Berks, B.C. (2014) Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. *The Journal of Biological Chemistry*, 289, 30889–30899. Available from: <https://doi.org/10.1074/jbc.M114.604892>
- Rodriguez, H., Fraga, R., Gonzalez, T. & Bashan, Y. (2006) Genetics of phosphate solubilization and its potential applications for improving plant growth-promoting bacteria. *Plant and Soil*, 287, 15–21. Available from: [https://doi.org/10.1007/978-1-4020-5765-6\\_2](https://doi.org/10.1007/978-1-4020-5765-6_2)
- Rossolini, G.M., Schippa, S., Riccio, M.L., Berlutti, F., Macaskie, L. E. & Thaller, M.C. (1998) Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cellular and Molecular Life Sciences*, 54, 833–850. Available from: <https://doi.org/10.1007/s000180050212>
- Santos-Beneit, F. (2015) The pho regulon: A huge regulatory network in bacteria. *Frontiers in Microbiology*, 6, 402. Available from: <https://doi.org/10.3389/fmicb.2015.00402>
- Saxena, A.K., Kumar, M., Chakdar, H., Anuroopa, N. & Bagyaraj, D.J. (2020) *Bacillus* species in soil as a natural resource for plant health and nutrition. *Journal of Applied Microbiology*, 128(6), 1583–1594. Available from: <https://doi.org/10.1111/jam.14506>
- Sebastian, M. & Ammerman, J.W. (2011) Role of the phosphatase PhoX in the phosphorus metabolism of the marine bacterium *Ruegeria pomeroyi* DSS-3. *Environmental Microbiology Reports*, 3, 535–542. Available from: <https://doi.org/10.1111/j.1758-2229.2011.00253.x>
- Spohn, M. & Kuzyakov, Y. (2013) Phosphorus mineralization can be driven by microbial need for carbon. *Soil Biology and Biochemistry*, 61, 69–75. Available from: <https://doi.org/10.1016/j.soilbio.2013.02.013>
- Srivastava, A., Saavedra, D.E.M., Thomson, B., Garcia, J., Zhao, Z., Patrick, W. et al. (2021) Enzyme promiscuity in natural environments: alkaline phosphatase in the ocean. *The ISME Journal*, 15, 3375–3383. Available from: <https://doi.org/10.1038/s41396-021-01013-w>
- Stopnisek, N., Bodenhausen, N., Frey, B., Fierer, N., Eberl, L. & Weissskopf, L. (2014) Genus-wide acid tolerance accounts for the biogeographical distribution of soil *Burkholderia* populations. *Environmental Microbiology*, 16(6), 1503–1512. Available from: <https://doi.org/10.1111/1462-2920.12211>
- Tan, H., Barret, M., Mooij, M.J., Rice, O., Morrissey, J., Dobson, A. 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, 661–672. Available from: <https://doi.org/10.1007/s00374-012-0755-5>
- Tang, J., Leung, A., Leung, C. & Lim, B.L. (2006) Hydrolysis of precipitated phytate by three distinct families of phytases. *Soil Biology & Biochemistry*, 38, 1316–1324. Available from: <https://doi.org/10.1016/j.soilbio.2005.08.021>
- Thaller, M.C., Schippa, S. & Rossolini, G.M. (1998) Conserved sequence motifs among bacterial, eukaryotic and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Science*, 1998(7), 1647–1652. Available from: <https://doi.org/10.1002/pro.5560070722>
- Thines, M., Aoki, T., Crous, P.W., Hyde, K., Lücking, R., Malosso, E. et al. (2020) Setting scientific names at all taxonomic ranks in

- italics facilitates their quick recognition in scientific papers. *IMA Fungus*, 11, 25. Available from: <https://doi.org/10.1186/s43008-020-00048-6>
- Tian, H., Chen, G., Zhang, C., Melillo, J.M. & Hall, C.A.S. (2010) Pattern and variation of C: N: P ratios in China's soils: a synthesis of observational data. *Biogeochemistry*, 98(1–3), 139–151. Available from: <https://doi.org/10.1007/s10533-009-9382-0>
- Tian, J., Ge, F., Zhang, D., Deng, S. & Liu, X. (2021) Roles of Phosphate Solubilizing Microorganisms from Managing Soil Phosphorus Deficiency to Mediating Biogeochemical P Cycle. *Biology*, 10(2), 158. Available from: <https://doi.org/10.3390/biology10020158>
- Tu, S.M.L., Ma, L. & Rathinasabapathi, B. (2011) Characterization of phytase from three ferns with differing arsenic tolerance. *Plant Physiology and Biochemistry*, 49, 146–150. Available from: <https://doi.org/10.1016/j.plaphy.2010.11.004>
- Udaondo, Z., Duque, E., Daddaoua, A., Caselles, C., Roca, A., Pizarro-Tobias, P. et al. (2020) Developing robust protein analysis profiles to identify bacterial acid phosphatases in genomes and metagenomic libraries. *Environmental Microbiology*, 22(8), 3561–3571. Available from: <https://doi.org/10.1111/1462-2920.15138>
- Wei, X., Hu, Y., Cai, G., Yao, H., Ye, J., Sun, Q. et al. (2021) Organic phosphorus availability shapes the diversity of phoD-harboring bacteria in agricultural soil. *Soil Biology and Biochemistry*, 161, 108364. Available from: <https://doi.org/10.1016/j.soilbio.2021.108364>
- Weon, H., Anandham, R., Kim, B., Hong, S., Jeon, Y. & Kwon, S. (2009) *Dyella soli* sp. nov. and *Dyella terrae* sp. nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, 59, 1685–1690. Available from: <https://doi.org/10.1099/ijs.0.004838-0>
- White, R.P., Murray, S., Rohweder, M., Prince, S. & Thompson, K. (2000) *Grassland ecosystems*. World Resources Institute: Washington DC, U.S.A.
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. New York: Springer Verlag.
- Wilke, A., Bischof, J., Harrison, T., Brettin, T., D'Souza, M., Gerlach, W. et al. (2015) A RESTful API for accessing microbial community data for MG-RAST. *PLoS Computational Biology*, 11(1), e1004008. Available from: <https://doi.org/10.1371/journal.pcbi.1004008>
- Yao, M.Z., Zhang, Y.H., Lu, W.L., Hu, M.Q., Wang, W. & Liang, A.H. (2012) Phytases: crystal structures, protein engineering and potential biotechnological applications. *Journal of Applied Microbiology*, 112, 1–14. Available from: <https://doi.org/10.1111/j.1365-2672.2011.05181.x>
- Zeng, J., Tu, Q., Yu, X., Qian, L., Wang, C., Shu, L. et al. (2022) PCycDB: a comprehensive and accurate database for fast analysis of phosphorus cycling genes. *Microbiome*, 10, 101. Available from: <https://doi.org/10.1186/s40168-022-01292-1>
- Zhou, Q., Daryanto, S., Xin, Z., Liu, Z., Liu, M., Cui, X. et al. (2017) Soil phosphorus budget in global grasslands and implications for management. *Journal of Arid Environments*, 144, 224–235. Available from: <https://doi.org/10.1016/j.jaridenv.2017.04.008>

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Garaycochea, S., Altier, N.A., Leoni, C., Neal, A.L. & Romero, H. (2023) Abundance and phylogenetic distribution of eight key enzymes of the phosphorus biogeochemical cycle in grassland soils. *Environmental Microbiology Reports*, 15(5), 352–369. Available from: <https://doi.org/10.1111/1758-2229.13159>