Rothamsted Repository Download

A - Papers appearing in refereed journals

Garaycochea, S., Altier, N., Leoni, C., Neal, A. L. and Romero, H. 2022. Abundance and phylogenetic distribution of eight key enzymes of the phosphorus biogeochemical cycle in grassland soils. *Environmental Microbiology.* https://doi.org/10.1111/1758-2229.13159ThisisanopenaccessarticleunderthetermsoftheCreativeCommonsAttributionLicense,whichpermitsuse,distributionandreproductioninanymedium,providedtheoriginalworkisproperlycited.©2023TheAuthors.EnvironmentalMicrobiologyReportspublishedbyAppliedMicrobiologyInternationalandJohnWiley&SonsLtd.EnvironmentalMicrobiologyReports.2023;1–18.wileyonlinelibrary.com/journal/emi41

The publisher's version can be accessed at:

- https://doi.org/10.1111/1758-2229.13159ThisisanopenaccessarticleunderthetermsoftheCreativeCommonsAttributionLicense,whichpermitsuse,distributionandreproductioninanymedium,providedtheoriginalworkisproperlycited.©2023TheAuthors.EnvironmentalMicrobiologyReportspublishedbyAppliedMicrobiologyInternationalandJohnWiley&SonsLtd.EnvironmentalMicrobiologyReports.2023;1–18.wileyonlinelibrary.com/journal/emi41

The output can be accessed at:

https://repository.rothamsted.ac.uk/item/989x3/abundance-and-phylogenetic-distribution-of-eight-key-enzymes-of-the-phosphorus-biogeochemical-cycle-in-grassland-soils.

© 10 May 2023, Please contact library@rothamsted.ac.uk for copyright queries.

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

Journal:	*Environmental Microbiology*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Garaycochea, Silvia; National Institute of Agricultural Research Experimental Station Wilson Ferreira Aldunate, Biotechnology Unit Altier, Nora; National Institute of Agricultural Research Experimental Station Wilson Ferreira Aldunate Leoni, Carolina; National Institute of Agricultural Research Experimental Station Wilson Ferreira Aldunate Neal, Andrew; Rothamsted at North Wyke Romero, Héctor; University of the Republic Faculty of Sciences, Laboratorio de Organización y Evolución del Genoma/Genómica Evolutiva, Departamento de Ecología y Evolución,; University of the Republic Faculty of Sciences, Laboratorio de Organización y Evolución del Genoma/Genómica Evolutiva, Departamento de Ecología y Evolución,
Keywords:	functional diversity, metagenomics/community genomics, soil microbes, microbiome
Abstract:	Grasslands are one of the most diverse and widely distributed biomes on the Earth's surface. Nutrient cycling is one of the main ecosystem services provided by grasslands. The organic fraction of phosphorus (P) represents over half of the total P in soil and is a valuable reservoir. Soil microorganisms, involved in the P cycle, mediate organic P release through three enzyme families: alkaline phosphatases, nonspecific acid phosphatases, and phytases (P-enzymes). This study aimed, through a metagenomic approach, to assess the abundance and phylogenetic distribution of prokaryotic P-enzymes in a wide distribution of grass biomes across the globe and how they are related with environmental variables. To generate a functional perspective of phosphorus cycling, 74 soil metagenomes from 17 sites/projects representing different environmental conditions were examined for eight key P-enzymes. Multivariate analyses showed that Tmax, pH and evapotranspiration were highly associated with P-enzymes abundance and diversity. In addition, they tend to respond in a correlated manner to these variables suggesting an intricate relationship of abundance and diversity between them. On the other hand, their association with the general functional profiles was more idiosyncratic.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

1 Title: Abundance and phylogenetic distribution of eight key enzymes of the phosphorus biogeochemical cycle in grassland

2 soils.

3

4 Running title: Prokaryotic P - enzymes in grassland soil.

5

6

7 Authors: Garaycochea S.*[1] , Altier N.[1] , Leoni C.[1] , Andrew L. Neal[3], Romero H.*[2]

8 [1] Instituto Nacional de Investigación Agropecuaria (INIA), Estación Experimental INIA Las Brujas, Ruta 48 Km 10,

9 Canelones, 90200, Uruguay.

10 [2] Laboratorio de Organización y Evolución del Genoma/Genómica Evolutiva, Departamento de Ecología y Evolución,

11 Facultad de Ciencias/CURE, Universidad de la República, Maldonado, Uruguay

12 [3] Net-Zero and Resilient Farming, Rothamsted Research, North Wyke, Okehampton, EX20 2SB, UK.

13 corresponding authors: Héctor Romero – eletor@fcien.edu.uy; Silvia Garaycochea – sgaraycochea@inia.org.uy

14

15

16 Silvia Garaycochea – sgaraycochea@inia.org.uy. Orcid: https://orcid.org/0000-0001-9685-0584

17 Nora Altier – naltier@inia.org.uy. Orcid: http://orcid.org/0000-0002-3173-6016

18 Carolina Leoni – cleoni@inia.org.uy. Orcid: http://orcid.org/0000-0002-3891-564X

19 Andrew L. Neal – andy.neal@rothamsted.ac.uk. Orcid: http://orcid.org/0000-0003-4225-1396

20 Héctor Romero – eletor@fcien.edu.uy. Orcid: https://orcid.org/0000-0001-9685-0584

21

25 Conflict of interest. None declared.

26

27

28

29

30 Abstract:

31

32 Grasslands are one of the most diverse and widely distributed biomes on the Earth's surface. Nutrient cycling is one of the

33 main ecosystem services provided by grasslands. The organic fraction of phosphorus (P) represents over half of the total P

34 in soil and is a valuable reservoir. Soil microorganisms, involved in the P cycle, mediate organic P release through three

35 enzyme families: alkaline phosphatases, nonspecific acid phosphatases, and phytases (P-enzymes). This study aimed,

36 through a metagenomic approach, to assess the abundance and phylogenetic distribution of prokaryotic P-enzymes in a wide

37 distribution of grass biomes across the globe and how they are related with environmental variables. To generate a

38 functional perspective of phosphorus cycling, 74 soil metagenomes from 17 sites/projects representing different

39 environmental conditions were examined for eight key P-enzymes. Multivariate analyses showed that Tmax, pH and

40 evapotranspiration were highly associated with P-enzymes abundance and diversity. In addition, they tend to respond in a

41 correlated manner to these variables suggesting an intricate relationship of abundance and diversity between them. On the

42 other hand, their association with the general functional profiles was more idiosyncratic.

43

44 **Keywords:** alkaline phosphatases, acid phosphatases, phytases, grasslands, phosphorus cycle

45

46

47

48

49

50

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 (Knapp et al., 2002; Lenhart et al., 2015; Barnet and Facey, 2016). 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 grassland biomass is above-ground: this, as well as the low rates of decomposition, generates significant accumulations of organic matter in grassland soils (Blair et al., 2014). Grasslands provide several key ecosystem services, such as food, fiber and forage production, water and nutrient cycling, and erosion control. Grassland biomes are habitat for a high diversity of plants, animals and microorganisms (Le Roux et al., 2011; Blair et al., 2014).

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). The principal C pool is soil organic matter, related to the type of vegetation and the amount of biomass below-ground. The availability of and cycling of the three nutrients are connected. N addition is known to influence multiple ecosystem functions (Li et al., 2015), including plant productivity, nutrient pools, organic matter decomposition and soil carbon stocks. Moreover, excessive N fertilization leads to an imbalance of N and P availability (Elser et al. 2009; Peñuelas et al. 2013; Cui et al., 2022). Soil P is present as two fractions, organic and inorganic phosphates. Low levels of soluble forms of P result from the high reactivity of both fractions with calcium (Ca), iron (Fe) or aluminum (Al) ions present in soil, forming complex insoluble associations (Achat et al, 2016).

The proportion of each P fraction varies between soils, dependent mainly upon the parent material, pH, temperature and organic matter inputs (Zhou et al., 2017; Gaiero et al., 2020). The organic fraction accounts for over half of total soil P and is a valuable reservoir that could be potentially mobilized (Condron et al., 2005; Haygarth et al., 2013; George et al., 2018). The organic P forms more abundant in soils are inositol phosphate, phospholipids, and nucleic acids (Gyaneshwar et al., 2002). Inositol phosphate (commonly called phytic acid) can account for up to 80% or more of total organic P (Quiquampoix and Mousain, 2005; Gerke, 2015). Phytic acid reacts with ions present in the soil forming stable and

79 insoluble complexes and so tends to accumulate in natural grasslands soils. Phospholipids and nucleic acids are both labile

80 and readily accessible to soil organisms (Gerke, 2015).

81

82 Soil microorganisms are an integral part of the soil P cycle, mediating P release for plants (Awasthi et al., 2011; Richardson

83 and Simpson, 2011). Mineralization of organic P by microbes is strongly influenced by environmental variables (Seshachala

84 and Tallapragada, 2012; Alori et al., 2017). Release of phosphate from organic P is mediated by three broad groups of

85 enzymes: 1) alkaline phosphatases that catalyze the hydrolysis between the carbon and phosphorus in organic phosphates

86 esters, 2) non-specific acid phosphatases, which perform the desphosphorylation of phospho-ester bonds or

87 phosphoanhydride in organic matter and 3) phytases, which specifically cause the release of P from phytic acid (Rossolini et

88 al., 1998; Huang et al., 2009; Jorquera et al., 2008; Bergkemper et al., 2016; Morrison et al., 2016; Gaiero et al., 2020).

89 Alkaline phosphatases are produced by a broad range of bacteria, archaea and fungi and are considered the most important

90 in the microbial P turnover (Li et al., 2021). They are grouped in three families, PhoA, PhoD and PhoX, with PhoD the most

91 abundant and ubiquitous (Ragot et al, 2015). Both PhoD and PhoX were identified as Ca^{2+}-dependent extracellular enzymes

92 and PhoA as a Zn^{2+}-dependent intracellular enzyme (Neal et al., 2018). All are regulated by P availability under control of

93 the Pho regulon (Santos-Beneit et al., 2015; Li et al., 2021). Alkaline phosphatases show a broad substrate specificity and

94 high catalytic efficiency (Rodriguez et al., 2014; Cai et al., 2021). These characteristics enable microorganisms harboring

95 these genes to use alternative P sources under P-limited conditions, conferring them an advantage over the plants (Li et al.,

96 2021).

97

98 Acid phosphatases are another group of enzymes distributed widely among microorganisms and plants. They are divided

99 into three families, Nsap class A, Nsap class B and Nsap class C, none of them exhibit strong substrate specificity, hence

100 their names (Thaller et al., 1998). These enzymes are mostly produced by microorganisms and are mostly active in acid

101 soils (Gaiero et al., 2018). Until now, the mechanisms involved in catalytic activity and its regulation are not well

102 understood. To expand the knowledge of these enzymes, metagenomic studies have been carried out to understand how they

103 vary in abundance and diversity in different environments (Bergkemper et al., 2016; Neal et al., 2018). Neal et al. (2018)

104 showed that Nsap class C, a putative extracellular enzyme, was predominant in acid soils under P limiting conditions

105 compared to Nsap class A, an intracellular enzyme mainly produced by plants and rhizosphere microorganisms. There is no

106 evidence that non-specific acid phosphatases are regulated by the Pho regulon.

107

108 Phytases are produced by bacteria, fungi, plants and animals able to catalyze mineralization of organic P from phytate to

109 inorganic P (Jorquera et al., 2008; Tu et al., 2011; Ariza et al., 2013). Phytase families more common in microorganisms are

110 the beta-propeller phytase (BPP), protein tyrosine phosphatase-like cysteine phytase (CPhy) and histidine acid phytase

111 (HAPhy) (Lim et al., 2007). The main differences between the phytase families are structural, mainly related to differences

112 in the active site which determines which phosphate group of the phytate is dephosphorylated, and co-factor requirements.

113 Despite this, all phytases can release the six phosphate molecules contained in phytate (Misset, 2002). The phytase enzymes

114 are distributed throughout soils, but the higher concentrations are found in the rhizosphere (Li et al., 2008). Phytases exhibit

115 different pH and temperature optima in the laboratory (Caffaro et al., 2020). Moreover, enzymatic activity is affected by soil

116 type, texture and mineralogy by varying the ability to retain an active enzyme (Rao et al., 1994; Tang et al., 2006; Azeem et

117 al., 2014)

118

119 The P cycle is determined by different environmental variables such as plant community, organic matter content, soil

120 properties and climatic conditions. Knowledge about the key microbial enzymes involved in organic P mineralization is

121 different to each class of enzymes. This study aimed, through a metagenomic approach, to assess for the first time how key

122 P cycling enzymes vary in their abundance and diversity and how this is related with environmental variables in the

123 grassland biome on a global scale. Grasslands, one of the five most important biomes on earth due to the biodiversity it

124 harbors and its economic importance, makes it necessary to have a deeper understanding of its functions and dynamics for

125 its preservation.

126

127

128 Experimental Procedures

129

130 Data collection

131 A total of 376 geo-referenced metagenomes samples from 17 projects deposited with MG-RAST were selected through the

132 TerrestrialMetagenomeDB (https://webapp.ufz.de/tmdb/) applying the following filters: Source DB: MG-RAST;

133 seq_technology: Ilumina; material: soil; Biome: grasslands, temperate grasslands, savanna and shrubland to assembly the

134 grassland soil metagenomes samples set. The set of environmental variables was assembled, including soil properties and

135 climatic variables for each sample based on its geographic location. Soil type and physicochemical properties were

136 obtained from SoilGrid 250m 2.0 – ISRIC World Soil Information. The following properties were included Bulk Density

137 (BD; cg cm^{-3}), Clay (g kg^{-1}), Sand (g kg^{-1}), Silt (g kg^{-1}), Cation Exchange Capacity at pH 7 (CEC; mmol(c) kg^{-1}), Total

138 Nitrogen (N; cg kg^{-1}), Soil Organic Carbon (SOC; cg kg^{-1}), pH (water*10). The organic available P (Pav) was estimated

139 based on SOC and N content following a model proposed by Tian et al. (2010). The climate variables were obtained from

140 TerraClimate   (https://www.climatologylab.org/terraclimate.html),   including maximum temperature, (T$_{max}$; ºC),

141 Precipitation, (ppt; mm), actual evapotranspiration (aet; mm), soil moisture (moisture; mm) and runoff (q; mm) (Table S1).

142 Hereafter they are called environmental variables. The collinearity analysis on the environmental variables set was

143 performed with R-base (R core Team 2022). The variables included in the set were those with r <= 0.5 and meaningful to

144 the study.

145

146

147 The soil metagenome sequencing from projects mgp91922 and mgp93346 (Uruguay) were carried out on a HiSeq Illumina

148 platform, (Service CD Genomics, NY, USA; pair-end read 150 bp). Raw sequence quality was analyzed with FastQC

149 software version 0.11.2. Assembly and functional annotation was performed on the MG-RAST repository. Raw sequence

150 data are publicly available on the MG-RAST repository.

151 The functional annotation based on MG-RAST subsystems level 2 of the 376 selected metagenomes was obtained from

152 MG-RAST repository (Meyer et al. 2008).

153 The set of predicted proteins to each metagenome was obtained through the RESTful API of MG-RAST (Wilke et al.,

154 2015). Protein sequences were downloaded using a matR version 0.9.1 package R (Braithwaite et al., 2018).

155

156 Statistical analyses

157

158 Uniform Manifold Approximation and Projection version 3 (UMAP, McInnes et al., 2018) was used to explore the

159 variability of metagenome functional profiles (subsystems level 2) implemented in the umap R package using default

160 parameters. Data normalization was done with the Variance Stabilizing Transformation (VST) method implemented in

161 DESeq2 R package (Love et al., 2014) with a nonspecific model.

162 Canonical analysis of principal coordinates (CAP) (Anderson and Willis 2003) implemented in Vegan R Package version

163 2.6.2 (Oksanen et al., 2019) was performed based upon Mahalanobis distance to calculate the relationship between the

164 metagenomes functional profiles (subsystems level 2) and environmental variables. Significance of the model parameters

165 was determined with permutational multivariate analysis of variance (ANOVA) with 999 permutations.

166 Hereafter, all analyses were performed on a reduced samples subset, in order to minimize the bias generated by the different

167 number of samples in each MG-RAST project. This subset included a maximum of 3 samples per MG-RAST project with

168 the same geo-reference. Finally, the reduced subset was made up of 74 samples from the 17 MG-RAST projects (Table S3).

169

170 Protein/function count matrix (level 4 in the MG-RAST nomenclature), including the 8 P-enzymes, for the 74 selected

171 metagenomes was normalized with CPM and TMM methods using the edgeR package (doi:

172 10.1093/bioinformatics/btp616). This data was used to perform the direct correlations of P-enzymes with environmental

173 variables.

174

175 The reference databases of the phosphatases and phytases (from now called P-enzyme) used in this work were built by Neal

176 et al., (2017). The P-enzymes included are listed in Table 1.

177

178 **Table 1**: List of P-enzymes included in the analyses.

Enzyme	Gene	Predicted Cellular Localization	Number of protein sequences in reference database
PhoA	*phoA*	Periplasmic/Cytoplasmic	293
PhoD	*phoD*	Outer membrane/extracellular	833
PhoX	*phoX*	Outer membrane/ extracellular	424
Nsap class A (Nsap-A)	*phoC*	Periplasmic/Cytoplasmic	750
Nsap class B (Nsap-B)	*aphA*	Periplasmic/Cytoplasmic	388
Nsap class C (Nsap-C)	*olpA*	Outer membrane/ extracellular	1123
β-propeller phytase (BPP)	*phyL, phyS*	Outer membrane/ extracellular	108
Cysteine phytase (Cphy)	*phyA*	Outer membrane/	122

extracellular

179

180

181 Protein sequence alignments of the respective reference database was performed using MAFFT version 7.4.60 (Katoh et al.,

182 2002) under default parameters. Reference protein phylograms were visualized with IQTree 2 version 1.6.12 (Minh et al.,

183 2020) and the evolutionary models were evaluated with RAxML-NG (Kozlov et al., 2019). Phylograms were plotted with

184 iTOL (Interactive Tree of Life; Letunic and Bork 2007).

185 To determine the abundance and diversity of the P-enzymes in the metagenomes, the predicted protein sets were aligned

186 with each P-enzyme reference database. First, the whole protein set was aligned against each P-enzyme reference database

187 using HMMER version 3.3.1 (hmmer.org). Then, the sequences with positive hits were extracted using the esl-fetch tool.

188 Finally, these sequences were aligned to the correspondent reference database using MAFFT with default parameters.

189 Maximum likelihood-based phylogenetic placement of metagenome-derived protein sequences on the appropriate P-enzyme

190 reference phylogram was performed with EPA-ng (Barbera et al., 2019). Edge-PCA ordination and Kantorovich-Rubinstein

191 (KR) distance metrics (Evans and Matsen, 2012; Matsen and Evans, 2013) were computed on these results. The edge-PCA

192 and KR distances were performed using gappa (Czech et al.,2020) and tree and domain composition diagrams were drawn

193 using Archaeopteryx (https://sites.google.com/site/cmzmasek/home/software/forester).

194

195 Canonical analysis of principal coordinates (CAP) (Anderson and Willis 2003) implemented in the Vegan R Package

196 version 2.6.2 (Oksanen et al., 2019) was used to evaluate the relationship between the abundance and diversity of P specific

197 functions with the environmental variables. A CAP analysis associating enzyme abundance with environmental variables

198 was performed using Mahalanobis distance. Each enzyme abundance in each sample was normalized in relation to the

199 sequencing coverage of each enzyme. Significance of the model parameters was determined with permutational multivariate

200 analysis of variance (PERMANOVA) with 999 permutations. The KR distance of each enzyme calculated as mentioned

201 above was used to perform the CAP analyses between abundance and diversity of each P-enzyme and environmental

202 variables. Significance of the model parameters was determined with PERMANOVA based upon 999 permutations.

203 Graphics were produced with the R package ggplot2 (Wickham, 2016). All basic statistical procedures were performed

204 using R-base (R core Team 2022).

205

206

Results

To generate a general perspective of grassland functional landscapes, an analysis was performed on a set of 376 grassland soil metagenomes. Based upon the distribution of 168 functional process or subsystem MG-RAST annotations (Table S1 and S2) we identified a strong association of clusters with geographic location/project and soil type: Kastazoem/Luvisol, Luvisol/Kastazoem, Fluvisol, Kastazoem (Fig. S1) based upon Uniform manifold approximation and projection (UMAP) analysis (McInnes et al., 2020). Besides this, samples from two different projects with soil classified as Phaeozem, were grouped together with Andosols. Two small groups of samples with soils classified as Kastazoem/Luvisol and Fluvisol were not clustered with soil samples with the same classification. Only one cluster was formed of samples from different projects and soil classifications. Within this cluster, a separation among samples was observed, mainly due to environmental variables. Samples from Brazil (Ferrasols), Uruguay (Vertisol/Phaeozem, Molisols/Phaeozem), and England (Cambisols), with high maximum temperatures (T_{max}) and soil moisture, and low pH clustered together. Samples from the USA (Chernozem and some Kastazoem) with low T_{max}, soil moisture and neutral pH form another cluster and two subgroups related to different soil types were also observed (Fig. S1). Subsequent analyses were performed on a balanced reduced subset of 74 grasslands soil metagenomes to minimize any bias associated with including different numbers of samples within each project (see methods and Table S3).

Metagenome subsystem profiles and environmental variables.

Canonical Analysis of Principal Coordinates (CAP) was performed to explore the relationship between environmental variables (Table S3) and soil microbial functional profiles (MG-RAST level 2 subsystem annotations, Table S4).The constrained model was significant (p = 0.001) and explained 24.8% of 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 moisture ($r = 0.536$). This axis separated samples from low pH soils (e.g.: mgp9904, mgp5588, mgp91922 and mgp93346) from those with neutral pH (mgp13948 among others). CAP2 axis was associated principally with pH ($r = -0.486$), SOC ($r = -0.220$), and T_{max} ($r = 0.664$), runoff (q, $r = 0.490$), 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}, moisture and ppt values of the set (Table 2, Table S5, Fig. S2).

235

Analyses based on abundance of P-enzymes genes.

Each metagenome predicted protein set was interrogated against of their correspondent reference database of the 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 abundances in each soil metagenome is shown in Table S6 and phylogenetic placements in Figure 1 and Figure S3.

The enzymes with highest relative abundance in global grasslands were alkaline phosphatases PhoD and PhoX (Table S6) both having broad phylogenetic distributions and no clear dominant phylotypes (Fig. 1a and Fig. 1b). PhoA had a median of 23.3 hits with a limited phylogenetic distribution (Fig. S3, Table S6).

Nsap-A and Nsap-C were the most abundant acid phosphatases and presented a different distribution in their corresponding phylogenetic trees. Whilst Nsap-A showed a broad distribution within its phylogeny (Fig. 1c), Nsap-C was linked to main branches of alpha and gammaproteobacteria, Flavobacteria, and Sphingobacteria classes. On the other side, Nsap-B had a low abundance and only gammaproteobacteria variants were found.

BPP was the most abundant phytase with a phylogenetic distribution mainly restricted to the Proteobacteria phylum (e.g.: *Pseudomonas*, *Alteromonas*, *Acinetobacter*) (Fig. 1d, Table S6). The CPhy phytase, with lower abundance, is distributed within Betaproteobacteria, Gammaproteobacteria and some classes of the Firmicutes phylum (Fig. S3, Table S6).

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 tree drawn with iTOL. The circle sizes represent the number of hits per node.

Correlation analyses between normalized enzyme abundance (by CPM and TMM methods) and environmental variables showed that PhoD, PhoX and BPP had a significant correlation with the pH, actual evapotranspiration (aet), precipitation (ppt), runoff (q) and soil moisture. In addition, PhoD showed significant correlation with SOC and Pav. NsapC showed significant correlation with aet, q, ppt and moisture (Table S7).

We used CAP analysis to explore the relationship between total enzyme abundance and environmental variables (Fig. S4). The constrained model based on Mahalanobis distance explained 36.4% of variance within the data set ($p = 0.001$). The

264 alkaline phosphatases PhoD, PhoX and PhoA were mainly associated with the explained variance. CAP1 axis explained

265 8.9% of variance ($p = 0.001$) and was associated with pH ($r = -0.70$), BD ($r = -0.48$), Sand content ($r = -0.40$), ppt ($r =$

266 0.74), aet ($r = 0.67$), SOC ($r = 0.52$), Pav ($r = 0.52$) and Silt ($r = 0.49$). Samples from metagenomes of clay soils with low

267 pH (5.0 to 6.8) and high SOC values (Androsols, Cambisols, Ferrasols, Fluvisols, Kastanozem/Luvisol,

268 Luvisol/Kastanozem) were separated in this axis from those of neutral or alkaline soils having lower SOC contents

269 (Chernozem, Luvisol and Kastanozem). CAP2 axis was associated with T_{max} ($r = -0.40$), BD ($r = -0.34$), pH ($r = -0.27$), ppt

270 ($r = 0.41$), actual evapotranspiration (aet, $r = 0.39$) and q ($r = 0.39$). Soil metagenomes associated with lower aet values and

271 relatively high T_{max} were separated on this axis. All variables were significant with a $p < 0.001$ (Table 2 and Table S8).

272

273 *Analyses based on the abundance and phylogeny of P-enzyme genes*

274 To gain deeper insight in the diversity and abundance of P-enzymes, CAP analyses were performed using Kantorovich –

275 Rubinstein (KR-CAP) distance matrices between samples to include not only abundance but also phylogenetic information.

276 PhoD KR-CAP analysis explained 49.8% of the total variance in the data set ($p < 0.001$). Eleven of the environmental

277 variables were associated with the first two KR-CAP axes. The KR-CAP1 axis was associated negatively with pH ($r = -$

278 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 ($r = 0.72$), Tmax

279 ($r = 0.313$), Silt ($r = 0.27$) and SOC ($r = 0.22$). This axis separated soils with low pH, relative high values of SOC and Tmax

280 (Cambisols, Ferrasols, Mollisols/Phazoem, Luvisol/Kastanozem) to the soils with higher pH and lower Tmax. The KR-

281 CAP2 axis was characterized by a negative association with T_{max} ($r = -0.41$), moisture ($r = -0.29$) and Silt ($r = -0.22$). This

282 axis separated soil with neutral pH and relatively high Silt and Sand values from the rest samples (Table S9, Figure 2).

283

284 **Figure 2:** CAP based on Kantorovich-Rubinstein distance for PhoD. PERMANOVA analysis with 999 permutations was

285 performed to determine the significance between the sites/MG-RAST project. For each MG-RAST project three samples

286 with the same geo-reference were included. Each point represents samples from the project mpg1992 (blue); mpg3520

287 (green); mpg5588 (dark red); mpg7792 (grey); mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mgp10523

288 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange);

289 mpg20922 (brown); mgp89409 (brick-red); mgp91922. (light green); mgp93346 (light blue). Vector lengths represent the

290 correlation between each variable and the axes. CAP analysis was performed with the Vegan R package and graphics were

291 produced with the R package ggplot2.

292

293   The same analysis was performed for the other enzymes and the results are summarized in Table 2 (Tables S10 to S16, Fig.

294   S5 to S7). Notably, pH, T_{max} and aet were associated with all enzyme distributions. SOC displayed a high relationship with

295   alkaline phosphatases PhoD and PhoX, and acid phosphatases Nsap-A and Nsap-C and estimated phosphorus (Pav) was

296   mainly associated with the phytases. Next, CEC showed a high correlation with alkaline phosphatases and phytases. Finally,

297   clay content was related mainly to alkaline phosphatases (Table 2).

298

299   **Table 2:** Canonical Correlation Analysis (CAP) summaries. CAP analyses based on Mahalanobis distance for MG-RAST

300   level 2 subsystem annotations and enzyme abundance and based on Kantorovich-Rubinstein distance for each of eight

301   enzymes. Environmental variables included: pH, Soil Organic Carbon (SOC), Phosphorous estimated (Pav), Cation

302   Exchange Capacity (CEC), Bulk Density (BD), Clay, Sand, silt, actual evapotranspiration (aet), runoff (q), soil moisture

303   (moisture), precipitation (ppt), maximum Temperature ($T_{max)}$. PERMANOVA analysis with 999 permutations was

304   performed to determine the significance between the sites/MG-RAST project. The values marked with red are significant.

305

306

307

308   *Covariation of P-enzymes genes*

309   In order to examine the co-variation between P-enzymes their KR-CAP analyses results were compared. First, the enzymes

310   were grouped in alkaline phosphatases (PhoD, PhoX and PhoA), acid phosphatases (NspA-to-C) and phytases (BPP and

311   Cphy), then a simple within group analysis was carried out by correlating their KR-CAP axes. For the case of the alkaline

312   phosphatases PhoD and PhoX there was a strong positive correlation between KR-CAP2 PhoD and KR-CAP1 PhoX (r =

313   0.87). Both PhoA KR-CAP axes showed significant correlations with KR-CAP2 PhoD (*r* = 0.41, *r* = 0.51, respectively) and

314   KR-CAP1 PhoX (*r* = 0.46 and *r* = 0.44, respectively) (Figure 3a, Table S17).

315   The acid phosphatases Nsap-A and Nsap-C KR-CAP axes were highly correlated (*r* = 0.78 for KR-CAP1s and *r* = 0.49 for

316   KR-CAP2). Moreover, both showed significant, albeit lower, correlations with Nsap-B (Figure 3b, Table S17).

317   Related to the phytases, the correlation analysis between KR-CAP axes of BPP and Cphy showed a significant and high

318   correlation between their CAP axes (Figure 3c, Table S17).

319   When comparing the most abundant enzymes between groups, a significative and high positive correlation among both

320   alkaline phosphatases PhoD and PhoX with the acid phosphatases Nsap-A and Nsap-C, was observed (Figure 3d, Table

321   S17)

322   Finally, when comparing each P-enzyme CAP analysis with the subsystem level CAP-analysis we observed that only PhoD

323   (CAP1-PhoD vs CAP2-SS = 0.44), BPP(CAP1-BPP vs CAP1-SS = -0.41, CAP2-BPP vs CAP2-SS = 0.55) and CPhy

324   (CAP1-Cphy vs CAP1-SS = -0.59), displayed significant correlations between the axes (Table S18).

325

326   **Figure 3:** Correlation matrix of KR-CAP axes. a) Correlogram of the alkaline phosphatases displays the pearson correlation

327   coefficients between KR-CAP PhoD axes, KR-CAP PhoX and KR-CAP PhoA; KR-CAP PhoX and KR-CAP PhoA. The

328   correlation coefficients are colored according to their values; being blue the positives values and red the negative values. b)

329   Correlogram of the acid phosphatases displays the Pearson correlation coefficients between KR-CAP Nsap-A axes, KR-

330   CAP Nsap-B and KR-CAP Nsap-C; KR-CAP Nsap-B and KR-CAP Nsap-C. The correlation coefficients are colored

331   according to their values; being blue the positives values and red the negative values. c) Correlogram of the phytases

332   displays the Pearson correlation coefficients between KR-CAP BPP axes and KR-CAP CPhy. The correlation coefficients

333   are colored according to their values; being blue the positives values and red the negative values. d) Correlogram of the

334   most abundance displays the Pearson correlation coefficients. The correlation coefficients are colored according to their

335   values; being blue the positives values and red the negative values. Correlation analysis and graphics were performed with

336   cor R package.

337

338   *Edge-PCA and taxonomic identification of differentially observed P-enzymes*

339   Edge-PCA analysis was applied to reveal the differential presence and abundance of P-enzyme variants among the soil

340   metagenomes; a summary of the results is shown in Table S19. Variability in the abundance of PhoD variants associated

341   with the first edge-PCA axis separated samples by soil type, pH and SOC content. The differences showed that the enzyme

342   variants of the species *Koribcater versatilis* (class Acidobacteriia) and *Rhodanobacter spathiphylli* (class

343   Gammaproteobateria) (Figure 4b) were more abundant in soils classified as Ferrasols, Cambisols, Molisols/Phaeozem and

344   Vertisol/Phaeozem with low pH and relatively high SOC content (left quadrant of Figure 4a and). Enzyme variants

345   associated with *Actinomyces*, *Bacillus* and *Planctomyces* (Figure 4b) were more abundant in Kastanozem, Chernozem,

346   Luvisol and Fluvisols soils with higher pH and lower SOC content (right quadrant of Figure 4a). The second axis was

347   associated with PhoD genes harbored by *Burkholderia* and *Acinetobacter* with higher abundance in soils with neutral pH

348   and low clay content (Tables S1 and S19, Figure 4a).

349

350   **Figure 4**: a) Graphic representation of the first two axes of the edge-PCA for PhoD using samples as observations. Each

351   point represents samples from the project mpg1992 (blue); mpg3520 (green); mpg5588 (dark red); mpg7792 (gray);

352   mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mgp10523 (stone blue); mgp10541 (turquoise); mgp10956

353   (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange); mpg20922 (brown); mgp89409 (brick-red); mgp91922

354   (light green); mgp93346 (light blue). b) The phylogeny distribution of PhoD hits along the first and second axis of the

355   analysis (protein with positive coefficients are marked in blue and proteins with negative coefficients are marked in orange).

356   The edge-PCA was performed using gappa software and tree and domain composition diagrams were drawn using

357   Archaeopteryx (https://sites.google.com/site/cmzmasek/home/software/forester).

358

359   The alkaline phosphatases PhoX and PhoA showed a narrower phylogenetic distribution and Alphaprotebacteria (*Rosevivax*

360   and *Agrobacterium* among others) genes were predominant in soils with high SOC values and relatively low T_{max} (Table S1

361   and Table S19). The *Burkholderia* variants were observed in soil samples with near neutral pH and average SOC and CEC

362   values (Fig. S8). The genes PhoA of *Pantoea* and *Providencia* together with *Acinetobacter* and *Actinobacter* genera were

363   associated with varying abundance between samples (Fig. S8). Again, *Acinetobater* was more abundant in soils with

364   *circum*-neutral pH and average SOC and CEC values (Fig. S8).

365   The acid phosphatases Nsap-A genes harbored by *Pedosphaera*, *Dyella jiangningensis*, *Dyella japonica* and *Rhodanobacter*

366   sp. were identified as the most abundant among soils with average SOC and CEC values and sandy texture (Fig. S8). On the

367   other hand, *Sphingomonas* sp., *Phenylobacterium* sp., *Rhodanobacter* sp., and *Caulobacter* species variants were more

368   abundant in soils with high clay content (Fig. S8).

369

370   The Nsap-C genes harboring by *Stenotrophomonas* (gamma-Proteobacteria) and *Podosphaera* were more abundant in soils

371   with low values of moisture and aet. *Enterobacter* Nsap-B variants predominated in soils classified as Ferrasols, Andosols

372   and Luvisol with acidic pH, low T_{max}, and high aet and ppt values. Metagenomes from Fluvisol, with a pH = 7, were

373   associated with a higher abundance of *Photobacter* and *Marinomonas* enzymes and were strikingly different to the rest

374   (Fig. S8).

1
2   375
3
4
5   376   BPP phytase genes were only found in soil samples with pH values above 6.6. BPP gene variants of the *Acinetobacter*,
6
7   377   *Pseudomonas*, *Methylophaga*, *Pseudoalteromonas* and *Alteromonadales* (Gammaprotobacteria), and *Shewanella* and
8
9   378   *Hylemonella* (Betaproteobacteria), were identified to dominate in clay soils with high CEC values. BPP genes harbored by
10
11 379   *Bacillus* species were most abundant in sandy soils with low nutrient content.
12
13 380
14
15 381   CPhy genes from *Bacillus*, *Acinteobacter* and *Pseudomonas* genera varied across the samples but there was no clear signal
16
17 382   to   reveal   associations   with   environmental   variables.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

383 Discussion

384

385 The ecosystems are connections among different factors such as soil types, plant communities, microbial communities,

386 macro and micro fauna, climate conditions among others (Islam et al., 2020). The present work aimed to study the microbial

387 functional diversity of P-enzymes in the grassland biome, using a metagenomic approach. The analyzed samples represent

388 different environmental conditions defined by the physical and chemical soil properties, and climate variables (Amundson,

389 2013; Islam et al. 2020). The effect of specific plant communities was not considered in the analyses and could constitute a

390 limitant in the study.

391

392 The UMAP analysis showed sample clustering mainly based on the sampling sites/projects with the exception of a

393 heterogeneous group which was represented by samples from Latin America, United States and England. As expected, these

394 samples came from soils with differences in their geological origin, nutritional status, and water content. This group enabled

395 the study of the relation among environmental variables and functional microbial profiles using multivariate methods. In

396 accordance with different authors microbial functional profiles were associated to T_{max} (Johnston et al., 2019; Sorensen et

397 al., 2019), soil moisture (Bachar et al., 2010; Ochoa et al., 2010) and pH (Fierer and Jackson., 2006; Islam et al., 2020) and

398 to a lesser extent with SOC (Lauber et al. 2013), soil texture (Tecon and Or, 2017; Garaycochea, et al. 2020) and

399 evapotranspiration (aet). So far, there are no reports about the relation between evapotranspiration and microbial functional

400 profiles. These results reveal the complexity of the relation between microbial taxonomy and functional diversity and the

401 constraints for the understanding of which factors influence the community assembly and functions.

402

403 Microbial enzymes such as phosphatases (Rodríguez et al. 2006, Nannipieri et al. 2011) and phytases (Yao et al. 2012; Tan

404 et al., 2013) have a crucial role in phosphorus cycling. This study showed that the environmental variables T_{max}, pH, SOC

405 and soil moisture are associated with alkaline phosphatase abundance as mentioned by Gu et al. (2020) and Gaiero et al.

406 (2020). Microbial alkaline phosphatases (PhoD, PhoX and PhoA) are considered the most important enzymes for P cycling

407 (Sphon and Kuzyakov, 2013). The present work found that PhoD was the most abundant and had the widest phylogenetic

408 distribution regardless of the soil properties such as pH. Bergkemper et al. (2016) reported that the *phoD* genes were

409 identified as the most abundant even in acidic soils and demonstrated the importance of alkaline phosphatases in the

410 mineralization of soil organic P. In relation to gene abundance, PhoD was followed by PhoX and PhoA which showed a

411 lower abundance and narrower phylogenetic distribution. Ragot et al (2015) and Neal et al (2017) also reported that PhoD

412 was the most abundant alkaline phosphatase with a cosmopolitan distribution in marine and terrestrial systems. On the other

413 hand, PhoX was reported as mainly abundant in terrestrial ecosystems (Ragot et al, 2017). PhoA has only been studied by

414 Sebastian and Ammerman (2011) in a marine ecosystem.

415

416 PhoD and PhoX showed a high correlation with SOC and clay. There are several studies which demonstrate the effect of

417 SOC, N and organic P content on the abundance and diversity of both enzymes and the corresponding bacteria (Ragot et al.,

418 2017; Wei et al., 2021; Li et al., 2021). The predicted extracellular location of both enzymes (Neal et al., 2017) may explain

419 the importance of clay content in relation to the stabilization role due to the immobilization and the maintenance of the

420 enzymatic activity (Margalef et al., 2017). Even though PhoD is widely distributed among different classes of bacteria, in

421 this study, variants associated with *Koribacter versatilis* (Acidobacteria class) and *Rhodanobacter spathiphylli*

422 (Gammaproteobacteria class) were the most abundant in soils with relatively high SOC values and low pH. The *phoD*

423 associated with *Koribater versatilis* has been identified as a dominant phylotype in a silty clay loam soil Chromic Luvisol in

424 the United Kingdom (Neal *et al*., 2017), and the *Rhodanobacter spathiphylli* associated phylotype has been identified as a

425 dominant phylotype in the rhizospheres of maize and sorghum in a Brazilian Distroferric Red Latosol (Neal *et al*., 2021).

426 Variants associated with *Actinomyces*, *Bacillus* and *Planctomyces* were prevalent in soils with lower SOC and neutral pH.

427 pH is the key variable explaining the difference in abundance of species since all reported species are heterotrophs (Kielak

428 et al., 2016; Saxena et al., 2020). Finally, the PhoX gene harbored by *Burkholdderia* genus differed in soils with low and

429 medium content of SOC. Bacteria from this genus show a wide repertoire of metabolic pathways making them more

430 competitive in nutrient-restrictive environments (Morya et al., 2020).

431

432 Considering acid phosphatases, Nsap-A and Nsap-C were the most abundant and highly correlated with pH, T_{max}, aet, SOC

433 and Sand content. Nsap-A was found in *Dyella jiangningensis*, *Dyella japonica* and *Rhodanobacter* sp. These species use

434 different carbon sources and have been reported to be dominant in acid and neutral soils (Weon et al., 2009; Dahal et al.,

435 2017). On the other hand, Nsap-C was identified in Alpha and Gammaproteobacteria, Flavobacteria, and Sphingobacteria

436 classes, consistent with previous evidence (Neal et al., 2017; Gaiero et al., 2020). Contrary to the finding that the two non-

437 specific acid phosphatases have similar abundance in a UK grassland soil (Neal et al., 2017), Nsap-A was the prevalent

438 phosphatase in the grasslands included in this study. The predominance of phosphatase on grassland could be affected by

439 the interaction between microorganism and plant communities since both are able to produce these enzymes (Mhlongp et

440 al., 2018).

441

442   The relationship study among each P-enzymes abundance and phylogeny with environmental variables demonstrated that

443   T_{max}, pH and aet were key players in the abundance and phylogenetic distribution of P-enzymes, as it was observed in the

444   analysis with the general functional profiles (i.e. subsystems level 2). More specifically, the co-variation analyses of P-

445   enzymes (Table S17) and of P-enzymes with the general functional profiles (Table S18) showed, on the one hand, that the

446   variation of P-enzymes present a highly interconnected structure between the 8 P-enzymes herein analyzed. On the other

447   hand, each of the P-enzymes showed a different relationship with the variability of the functional profile in each sample.

448

449   One important question is to understand if the P-enzymes are driven particularly by the change of certain organisms that are

450   carrying them or, in turn, they are following the general major changes in the community structure. One possible hint in this

451   direction is given by the comparison of CAP analyses at the subsystem level with the ones of each P-enzyme (Table S18).

452   Indeed, in this case only PhoD and the two phytases (BPP and Cphy) showed a similar trend in these analyses, indicating

453   that these enzymes may be accompanying the general change in the functional structure of the metagenome, whilst the other

454   P-enzymes are varying in a more idiosyncratic manner.

455

456   This work studied the abundance and distribution of eight key enzymes involved in P organic cycling. The environmental

457   variables explained a low proportion of the observed microbial functional diversity. However, T_{max}, pH and aet were related

458   to the diversity of almost all enzymes. These results are in relation to the geographical global scale of the study. The use of

459   information from samples from very distant sites determines only the effect on the diversity of the variables with greater

460   differences among the sites. Likewise, it was possible to identify the effect of other variables with a more localized effect,

461   such as soil texture and nutrient content, as important determinants of microbial community structure and functions. The

462   complexity of the studied system requires a combination of approaches and the generation of local data that allow the

463   understanding of factors affecting the presence of bacteria carrying P-enzymes genes as well as their functionality.

464

467

468

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. doi: https://doi.org/10.1007/s10533-015-0178-0.

Alori E.T., Glick, B.R., Babalola, O. (2017). Microbial Phosphorus Solubilization and Its Potential for Use in Sustainable Agriculture..Front. Microbiol. 8:971. doi: 10.3389/fmicb.2017.00971

Amundson, R. (2013). Soil formation. In: Holland HD, Turekian KK (eds) Treatise on geochemistry, 2nd edn. Elsevier Science, USA, pp 1–26. doi: 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. doi:

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., et al. (2013) Degradation of Phytate by the 6-Phytase from *Hafnia alvei*: A Combined Structural and Solution Study. PLoS ONE 8(5): e65062. doi: 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., Imran, M. (2015) Microbial phytase activity and their role in organic P mineralization, Archives of Agronomy and Soil Science, 61:6, 751-766. doi: 10.1080/03650340.2014.963796

492 Bachar, A., Al-Ashhab, A., Soares, M.I.M., Sklars, M., Angel, R., Ungar, R., Gillor, O. (2010). Soil microbial abundance
493 and diversity along a low precipitation gradient. Microb Ecol 60:453–461. doi: https://doi.org/10.1007/s00248-010-9727-1

494 Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., Stamatakis, A. (2019). EPA-ng: Massively Parallel
495 Evolutionary Placement of Genetic Sequences. Systematic Biology, 68, 365–369. doi:
496 https://doi.org/10.1093/sysbio/syy054

497 Barnett K.L., Facey S.L. (2016). Grasslands, Invertebrates, and Precipitation: A Review of the Effects of Climate Change.
498 Front. Plant Sci. 7:1196. doi: 10.3389/fpls.2016.01196

499 Bergkemper, F., Kublik, S., Lang, F., Krüger, J., Vestergaard, G., Schloter, M., Schulz, S. (2016). Novel oligonucleotide
500 primers reveal a high diversity of microbes which drive phosphorous turnover in soil. J. Microbiol. Methods 125, 91–97.
501 doi:https://doi.org/10.1016/j.mimet.2016.04.011.

502 Blair, J., Nippert, J., Briggs, J. (2014). In R.K. Monson (ed.). Grassland Ecology, Ecology and the Environment, 389-423.
503 10.1007/978-1-4614-7501-9_14.

504 Braithwaite, D.T., Keegan, K.P. (2018). matR: Metagenomics Analysis Tools. R package version 0.9.1.   https://CRAN.R-
505 project.org/package=matR.

506 Caffaro, M., Balestrasse, K., Rubio, G. (2020). Adsorption to soils and biochemical characterization of commercial
507 phytases. SOIL. 6. 153-162. doi:10.5194/soil-6-153-2020.

508 Cai, S., Deng, K., Tang, J., Sun, R., Lu, H., Li, et al. (2021). Characterization of extracellular phosphatase activities in
509 periphytic biofilm from paddy field. Pedosphere 31, 116–124. doi:https://doi.org/10.1016/S1002-0160(20)60061-3.

510 Condron, L.M., Turner, B.L., Cade-Menun, B.J. (2005). Chemistry and dynamics of soil organic phosphorus. Phosphorus
511 Agric. Environ. 46, 87–121. doi: https://doi.org/10.2134/agronmonogr46.c4.

512 Cui, H., Sun, W., Delgado-Baquerizo, M., Ma., J.Y., Wang, K., Ling, X. (2022). Phosphorus addition regulates the
513 responses of soil multifunctionality to nitrogen over-fertilization in a temperate grassland. Plant Soil **473**, 73–87. doi:
514 https://doi.org/10.1007/s11104-020-04620-2.

515 Czech, L., Barbera, P., Stamatakis, A. (2020). Genesis and Gappa: processing, analyzing and visualizing phylogenetic

516 (placement) data. Bioinformatics, 36, 3263–3265. doi:https://doi.org/10.1093/bioinformatics/btaa070.

517 Dahal, R.H., Kim, J. (2017). *Rhodanobacter humi sp.* nov., an acid-tolerant and alkalitolerant gammaproteobacterium

518 isolated from forest soil. Int J Syst Evol Microbiol, 67, 1185-1190. doi: 10.1099/ijsem.0.001786.

519 Dubeux, J., Sollenberger, L. E., Mathews, B. W. , Scholberg, J. M., Santos, H. Q. (2007). Nutrient Cycling in Warm-

520 Climate Grasslands. Crop Sci. 47:915–928. doi: 10.2135/cropsci2006.09.0581

521 Elser, J., Andersen, T., Baron, J.S., Bergström, AK., Jansson, M., Kyle, M., et al. (2009). Shifts in Lake N:P Stoichiometry

522 and Nutrient Limitation Driven by Atmospheric Nitrogen Deposition. Science , 326, N.º 5954, 835-837. doi:DOI:

523 10.1126/science.1176199.

524

525 Evans, S.N., Matsen, F.A. (2012). The phylogenetic Kantorovich–Rubinstein metric for environmental sequence samples. J

526 R Stat Soc Series B Stat Methodol, 74:569–92. doi: 10.1111/j.1467-9868.2011.01018.x

527 Fierer, N., Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci

528 103:626–631. doi: https://doi.org/10.1073/pnas.0507535103

529 Gaiero, J.R., Bent, E., Fraser, T.D., Condron, L.M., Dunfield, K.E. (2018). Validating novel oligonucleotide primers

530 targeting three classes of bacterial non-specific acid phosphatase genes in grassland soils. Plant Soil 427, 39–51. doi:

531 https://doi.org/10.1007/s11104-017-3338-2.

532 Gaiero, J., Bent, E., Boitt, G., Condron, L., Dunfield, K. (2020). Effect of long-term plant biomass management on

533 phosphatase-producing bacterial populations in soils under temperate grassland. Applied Soil Ecology, 151, 103583. doi:

534 https://doi.org/10.1016/j.apsoil.2020.103583.

535 Garaycochea, S., Romero, H., Beyhaut, E., Neal, A.L, Altier, N. (2020). Soil structure, nutrient status and water holding

536 capacity shape Uruguayan grassland prokaryotic communities. *FEMS Microbiology Ecology*, 96, 12. doi:

537 https://doi.org/10.1093/femsec/fiaa207

538 George, T.S., Giles, C., Menezes-Blackburn, D., Condron, L., Gama-Rodrigues, A.C., Jaisi, D., et al. (2018). Organic

539 phosphorus in the terrestrial environment: a perspective on the state of the art and future priorities. Plant Soil 427, 191–208.

540 doi: https://doi.org/10.1007/s11104-017-3391-x.

541 Gerke, J., (2015). The acquisition of phosphate by higher plants: Effect of carboxylate release by the roots. A critical review.

542 J. Plant Nutr. Soil Sci. 178, 351–364. doi: https://doi.org/10.1002/jpln.201400590.

543 Gu, C., Wilson, S. G., Margenot, A. (2020). Lithological and bioclimatic impacts on soil phosphatase activities in California

544 temperate forests. Soil Biology and Biochemistry. 141, 107633. doi:https://doi.org/10.1016/j.soilbio.2019.107633.

545 Gyaneshwar P, Kumar GN, Parekh LJ, Poole PS. (2002). Role of soil microorganisms in improving P nutrition of plants.

546 Plant and Soil. 245:83–93. doi: https://doi.org/10.1023/A:1020663916259

547

548 Haygarth, P.M., Bardgett, R.D., Condron, L.M. (2013). Nitrogen and phosphorus cycles and their management. In: Soil

549 Conditions and Plant Growth. Blackwell Publishing Ltd, Oxford, pp. 132–159. doi:

550 https://doi.org/10.1002/9781118337295.ch5.

551 Huang, H., Shi, P., Wang, Y., Luo, H., Shao, N., Wang, G., et al. (2009). Diversity of beta-propeller phytase genes in the

552 intestinal contents of grass carp provides insight into the release of major phosphorus from phytate in nature. Appl. Environ.

553 Microbiol. 75, 1508–1516. doi: https://doi.org/10.1128/AEM.02188-08.

554 Islam, W., Noman, A., Naveed, H. Huang, Z., Chen., H.. Role of environmental factors in shaping the soil microbiome.

555 *Environ Sci Pollut Res*. 27, 41225–41247 (2020). doi: https://doi.org/10.1007/s11356-020-10471-2.

556 Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., Van Veen, J.A., Kuramae, E.E. (2016). The Ecology of Acidobacteria:

557 Moving beyond Genes and Genomes. Front. Microbiol. 7:744. doi: 10.3389/fmicb.2016.00744.

558 Johnston, E.R., Hatt, J.K., He, Z., Wu, L., Guo, X., Luo, Y. et al. (2019). Responses of tundra soil microbial communities to

559 half a decade of experimental warming at two critical depths. Proc Natl Acad Sci USA . doi: 116:15096–15105.

560 https://doi.org/10.1073/pnas.1901307116

561 Jorquera, M., Martínez, O., Maruyama, F., Marschner, P., de la Luz Mora, M. (2008). Current and future biotechnological

562 applications of bacterial phytases and phytase producing bacteria. Microbes Environ. 23, 182–191. doi:

563 https://doi.org/10.1264/jsme2.23.182.

564 Katoh, K., Misawa, K., Kuma, K., Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based

565 on fast Fourier transform. Nucleic Acids Res. 15, 3059-66. doi: 10.1093/nar/gkf436.

566 Knapp, A. K., Harper, C. W., Danner, B. T., Lett, M. S. (2002). Rainfall variability, carbon cycling, and plant species

567 diversity in a mesic grassland. Science 298, 2202–2205. doi: 10.1126/science.1076347.

568 Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool

569 for maximum likelihood phylogenetic inference. Bioinformatics, 35, 4453–4455. doi:

570 https://doi.org/10.1093/bioinformatics/btz305

571

572 Lauber , C.L., Ramirez, K.S., Aanderud, Z., Lennon, J., Fierrer, N. (2013) Temporal variability in soil microbial

573 communities across land-use types. ISME J 7:1641–1650. https://doi.org/10.1038/ismej.2013.50

574 Le Roux, X., Recous, S., Attard, E. 2011. Soil microbial diversity in grasslands, and its importance for grassland

575 functioning and services. In G. Lemaire and A. Chabbi. (eds). Grassland Productivity and Ecosystem Services. CABI

576 International, Wallingford, UK, 158–165.

577 Lenhart, P. A., Eubanks, M. D., and Behmer, S. T. (2015). Water stress in grasslands: dynamic responses of plants and

578 insect herbivores. Oikos 124, 381–390. doi:10.1111/oik.01370.

579 Letunic, I., Bork, P. (2007). Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation.

580 Bioinformatics, 23, 127–8. doi: 10.1093/bioinformatics/btl529

581 Li, Y. F., Luo, A. C., Wei, X. H., Yao, X. G. (2008). Changes in phosphorus fractions, pH, and phosphatase activity in the

582 rhizosphere of two rice genotypes, Pedosphere, 18, 785–794. doi: https://doi.org/10.1016/S1002-0160(08)60074-0.

583 Li, Y., Niu, S., Yu, G. (2015). Aggravated phosphorus limitation on biomass production under increasing nitrogen loading:

584 a meta-analysis. Glob Chan Biol, 22, 934-943. doi: https://doi.org/10.1111/gcb.13125.

585   Li, J., Xie, T., Zhu, H., Zhou, J., Li, c., Xiong, W., et al. (2021). Alkaline phosphatase activity mediates soil organic

586   phosphorus   mineralization   in   a   subalpine   forest   ecosystem,   Geoderma,   404,   115376.   doi:

587   https://doi.org/10.1016/j.geoderma.2021.115376.

588   Lim, B.L., Yeung, P., Cheng, C., Hill, J.E. (2007). Distribution and diversity of phytate-mineralizing bacteria. ISME J

589   1:321–330. doi: https://doi.org/10.1038/ismej.2007.40.

590   Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.

591   (2014). Genome Biol, 15:550. doi: https://doi.org/10.1186/s13059-014-0550-8.

592

593   Margalef, O., Sardans, J., Fernández-Martínez, M., Molowny-Horas, R., Janssens, I. A., Ciais, P., et al. (2017). Global

594   patterns of phosphatase activity in natural soils. Sci Rep **7**, 1337. doi: https://doi.org/10.1038/s41598-017-01418-8.

595   Matsen, F.A., Evans, S.N. (2013). Edge principal components and squash clustering: using the special structure of

596   phylogenetic   placement   data   for   sample   comparison.   PLoS   One   ,   8:e56859.

597   doi:https://doi.org/10.1371/journal.pone.0056859

598   McInnes, L., Healy, J., Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension

599   Reduction. Stat and ML, arXiv:1802.03426v3. doi: https://doi.org/10.48550/arXiv.1802.03426.

600   Mhlongo, M.I., Piater, L.A., Madala, N.E., Labuschagne, N., Dubery, I.A. (2018). The Chemistry of Plant–Microbe

601   Interactions in the Rhizosphere and the Potential for Metabolomics to Reveal Signaling Related to Defense Priming and

602   Induced Systemic Resistance. Front. Plant Sci. 9:112. doi: 10.3389/fpls.2018.00112

603   Minh, B., Schmidt, H., Chernomor, O., Schrempf, D., Woodhams, M., Von Haeseler, A., Lanfear, R. (2020). IQ-TREE 2:

604   New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution, 37,

605   1530–1534. doi: https://doi.org/10.1093/molbev/msaa015.

606   Misset, O.: Phytase, in: Handbook of Food Enzymology, edited by: Whitaker, J., Voragen, A., and Wong, D., Boca Raton,

607   CRC Press, (2002) 687–706. doi:https://doi.org/10.1201/9780203910450.

608 Morrison, E., Newman, S., Bae, H.S., He, Z., Zhou, J., Reddy, K.R., Ogram, A. (2016). Microbial genetic and enzymatic

609 responses to an anthropogenic phosphorus gradient within a subtropical peatland. Geoderma 268, 119–127. doi:

610 https://doi.org/10.1016/j.geoderma.2016.01.008.

611 Morya, R., Salvachúa, D., Thakur, I. S. (2020). Burkholderia: An Untapped but Promising Bacterial Genus for the

612 Conversion of Aromatic Compounds. Trends in Biotechnol. 38, 9: 963-975. doi:

613 https://doi.org/10.1016/j.tibtech.2020.02.008

614 Nannipieri, P., Giagnoni, L., Landi, L., Renella, G. (2011). Role of phosphatase enzymes in soil. In: Phosphorus in Action.

615 Soil Biology Springer, Heidelberg, 215–243. doi:10.1007/978-3-642-15271-9_9

616

617 Neal, A.L., Rossmann, M., Brearley, C., Akkari, E., Guyomar, C., Clark, I., et al. (2017). Land-use influences phosphatase

618 gene microdiversity in soils. Environ Microbiol. 19:2740–753. doi:10.1111/1462-2920.13778

619 Neal, A. L., Blackwell, M. S. A., Akkari, E., Clark, I. M., Hirsch, P. R. and Guyomar, C. (2018). Phylogenetic distribution,

620 biogeography and the effects of land management upon bacterial non-specific Acid phosphatase Gene diversity and

621 abundance. Plant and Soil. 427, 175-189. doi:https://doi.org/10.1007/s11104-017-3301-2 .

622 Neal, A. L., McLaren, T., Lourenço Campolino, M., David, H., Marcos Coelho, A., Gomes de Paula Lana, U., et al. (2021).

623 Crop type exerts greater influence upon rhizosphere phosphohydrolase gene abundance and phylogenetic diversity than

624 phosphorus fertilization. FEMS Microbiol. Ecol., 97, 4. doi: https://doi.org/10.1093/femsec/fiab033

625 Ochoa,M.S., Pedraza, R.M., Martínez, M. (2010). Plantas, hongos micorrízicos y bacterias : su compleja red de

626 interacciones. Biológicas 12:65–71.

627 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. et al. Vegan: community ecology

628 package. R package version 2.62. https: //CRAN.R-project.org/package=vegan, 08/15/2022.

629 Peñuelas, J., Poulter, B., Sardans, J., Philippe, C., Van der Velde, M., Laurent, B. et al. (2013) Human-induced nitrogen–

630 phosphorus imbalances alter natural and managed ecosystems across the globe. Nat Commun **4**, 2934. doi:

631 https://doi.org/10.1038/ncomms3934.

632 Quiquampoix H, Mousain D. (2005). Enzymatic hydrolysis of organic phosphorus. In: Turner BL, Frossard E, Baldwin D,

633 (eds). Organic phosphorus in the environment. Wallingford (UK): CABI, 89–112. doi:

634 https://doi.org/10.1079/9780851998220.0089.

635 Ragot, S., Kertesz, M., Bünemann, E. (2015). phoD Alkaline Phosphatase Gene Diversity in Soil. Applied and

636 Environmental Microbiology. doi: 81. 10.1128/AEM.01823-15.

637 Ragot, S., Kertesz, M., Mészáros, E., Frossard, E., Bünemann, E. (2017). Soil *phoD* and *phoX* alkaline phosphatase gene

638 diversity responds to multiple environmental factors, FEMS Microbiology Ecology. 93, 1. doi:

639 https://doi.org/10.1093/femsec/fiw212

640 Rao, M.A., Violante, A., Gianfreda, L. (1994). Catalytic behavior of acid phosphatase immobilized on clay minerals and

641 organo-mineral complexes transactions. In: 15th World Congress of Soil Science; volumen 3b. International Soil Science

642 Society, Mexico. Acapulco (Mexico): Mexican Society of Soil Science; 117–118.

643 Richardson, A., Simpson,R. (2011). Soil microorganisms mediating phosphorus availability. Plant Physiol. 156: 989-996.

644 doi: 10.1104/pp.111.175448.

645 Rodriguez, H., Fraga, R., Gonzalez, T., Bashan, Y. (2006). Genetics of phosphate solubilization and its potential

646 applications for improving plant growth-promoting bacteria. Plant Soil 287, 15–21. doi: 10.1007/978-1-4020-5765-6_2

647 Rodriguez, F., Lillington, J., Johnson, S., Timmel, C. R., Lea, S. M., Berks, B. C. (2014). Crystal structure of the Bacillus

648 subtilis phosphodiesterase PhoD reveals an iron and calcium-containing active site. J Biol Chem, 289, 30889–30899. doi:

649 10.1074/jbc.M114.604892

650 Rossolini, G,M,, Schippa, S., Riccio, M.L.,Berlutti, F., Macaskie, L.E., Thaller, M.C. (1998). Bacterial nonspecific acid

651 phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. Cell Mol Life Sci. 54: 833-850. doi:

652 10.1007/s000180050212

653 Santos-Beneit, F. (2015). The Pho regulon: A huge regulatory network in bacteria. Front. Microbiol, 6, 402. doi:

654 10.3389/fmicb.2015.00402.

655 Saxena, A.K., Kumar, M., Chakdar, H., Anuroopa, N., Bagyaraj, D.J. (2020). Bacillus species in soil as a natural resource

656 for plant health and nutrition. J Appl Microbiol. 128(6):1583-1594. doi: 10.1111/jam.14506.

657 Sebastian, M., Ammerman, J.W. (2011). Role of the phosphatase PhoX in the phosphorus metabolism of the marine

658 bacterium *Ruegeria pomeroyi* DSS-3. Environ Microbiol Rep 3:535–542. doi: 10.1111/j.1758-2229.2011.00253.x.

659 Seshachala, U., Tallapragada, P. (2012). Phosphate solubilizers from the rhizosphere of Piper nigrum L. in Karnataka, India.

660 Chil. J. Agric. Res. 72, 397–403. doi: 10.4067/S0718-58392012000300014.

661

662 Sorensen, P.O., Bhatnagar, J.M., Christenson, L. Duran, J., Fahey, T., Fisk, M. et al .(2019). Roots mediate the effects of

663 snowpack decline on soil bacteria, fungi, and nitrogen cycling in a northern hardwood forest. Front Microbiol 10:926. doi:

664 https://doi.org/10.3389/fmicb.2019.00926.

665 Spohn, M., Kuzyakov, Y. (2013). Phosphorus mineralization can be driven by microbial need for carbon. Soil Biology and

666 Biochemistry, 61, 69–75. doi: https://doi.org/10.1016/J.SOILBIO.2013.02.013

667 Tan, H., Barret, M., Mooij, M.J., Rice, O., Morrissey, J., Dobson, A., *et al.* (2013). Long-term phosphorus fertilisation

668 increased the diversity of the total bacterial community and the *phoD* phosphorus mineraliser group in pasture soils. *Biol*

669 *Fertil Soils* **49**, 661–672. doi: https://doi.org/10.1007/s00374-012-0755-5.

670 Tang,J., Leung, A., Leung,C., Lim, B.L. (2006). Hydrolysis of precipitated phytate by three distinct families of phytases.

671 Soil Biol Chem. 38:1316–1324. doi: https://doi.org/10.1016/j.soilbio.2005.08.021.

672 Tecon, R., Or, D. (2017). Biophysical processes supporting the diversity of microbial life in soil. FEMS Microbiol Rev

673 41:599–623. doi:https://doi.org/10.1093/femsre/fux039

674 Thaller, M.C., Schippa,S., Rossolini, G.M. (1998). Conserved sequence motifs among bacterial, eukaryotic and archaeal

675 phosphatases that define a new phosphohydrolase superfamily. Protein Sci. 1998; 7: 1647-1652. doi:

676 10.1002/pro.5560070722.

677 Tu, S.M.L., Ma, L., Rathinasabapathi, B. (2011). Characterization of phytase from three ferns with differing arsenic

678 tolerance. Plant Physiol Biochem. 49:146–150. doi: https://doi.org/10.1016/j.plaphy.2010.11.004.

679   Wei, X., Hu, Y., Cai, G., Yao, H., Ye, J., Sun, Q., et al. (2021). Organic phosphorus availability shapes the diversity of

680   phoD-harboring bacteria in agricultural soil. Soil Biol Biochem. 161, 108364. doi:

681   https://doi.org/10.1016/j.soilbio.2021.108364.

682   Weon, H., Anandham, R., Kim, B., Hong, S., Jeon, Y., Kwon, S. (2009). *Dyella soli sp*. nov. and *Dyella terrae sp. nov.*,

683   isolated from soil. Int. J. Syst. Evol. Microbiol, 59, 1685–1690. doi: https://doi.org/10.1099/ijs.0.004838-0.

684

685   White, R.P., Murray, S., Rohweder, M., Prince, S., Thompson, K. (2000). Grassland ecosystems. World Resources Institute.

686   Washington DC, U.S.A.

687   Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer Verlag: New York. 2016.

688   Wilke, A., Bischof, J., Harrison, T., Brettin, T., D'Souza, M., Gerlach W, *et al*. (2015) A RESTful API for Accessing

689   Microbial   Community   Data   for   MG-RAST.   PLoS   Comput   Biol   11(1):   e1004008.

690   https://doi.org/10.1371/journal.pcbi.1004008

691   Yao, M.Z., Zhang, Y.H., Lu, W.L., Hu,M.Q., Wang, W., Liang, A.H. (2012). Phytases: crystal structures, protein

692   engineering and potential biotechnological applications. J Appl Microbiol 112:1–14. doi:10.1111/j.1365-2672.2011.05181.x

693   Zhou, Q., Daryanto, S., Xin, Z., Liu, Z., Liu, M., Cui, X., Wang, L. (2017). Soil phosphorus budget in global grasslands and

694   implications for management. Journal of Arid Environments. 144. 10.1016/j.jaridenv.2017.04.008.

695

696

697

698

699

700

701

702

703

704

705 Tables legends:

706

707 **Table 1**: List of P-enzymes included in the analyses.

708 **Table 2**: Canonical Correlation Analysis (CAP) summaries. CAP analyses based on Mahalanobis distance for MG-RAST

709 level 2 subsystem annotations and enzyme abundance and based on Kantorovich-Rubinstein distance for each of eight

710 enzymes. Environmental variables included: pH, Soil Organic Carbon (SOC), Phosphorous estimated (Pav), Cation

711 Exchange Capacity (CEC), Bulk Density (BD), Clay, Sand, silt, actual evapotranspiration (aet), runoff (q), soil moisture

712 (moisture), precipitation (ppt), maximum Temperature (T_{max}). PERMANOVA analysis with 999 permutations was

713 performed to determine the significance between the sites/MG-RAST project. The values marked with red are significant.

714

715 Figures legends:

716

717 **Figure 1:** Phylogenetic placements of the predicted proteins of each metagenome with respect to the reference bases of

718 each enzyme: a) PhoD, b) PhoX, c) Nsap-A and d) BPP The size of the circle representing placements is proportional to the

719 abundance. Maximum likelihood-based phylogenetic placement of metagenome-derived protein sequences was performed

720 with EPA-ng and tree drawn with iTOL. The circle sizes represent the number of hits per node.

721

722

723

724 **Figure 2:** CAP based on Kantorovich-Rubinstein distance for PhoD. PERMANOVA analysis with 999 permutations was

725 performed to determine the significance between the sites/MG-RAST project. For each MG-RAST project three samples

726 with the same geo-reference were included. Each point represents samples from the project mpg1992 (blue); mpg3520

727 (green); mpg5588 (dark red); mpg7792 (grey); mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mgp10523

728 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange);

729 mpg20922 (brown); mgp89409 (brick-red); mgp91922. (light green); mgp93346 (light blue). Vector lengths represent the

730 correlation between each variable and the axes. CAP analysis was performed with the Vegan R package and graphics were

731 produced with the R package ggplot2.

732

733

734

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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red the negative values. d) Correlogram of the most abundance displays the Pearson correlation coefficients. The correlation coefficients are colored according to their values; being blue the positives values and red the negative values. Correlation analysis and graphics were performed with cor R package.

746

747

748

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 (gray); mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mpg10523 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange); mpg20922 (brown); mgp89409 (brick-red); mgp91922 (light green); mgp93346 (light blue). b) The phylogeny distribution of PhoD hits along the first and second axis of the analysis (protein 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).

757

758

759

760

761

Hoja1

Table1: Canonical Correlation Analysis (CAP) summary.
Canonical Correlation Analysis (CAP) summaries. CAP analyses based on Mahalanobis distance for MG-R/
or each of eight enzymes. Environmental variables included: pH, Soil Organic Carbon (SOC), Phosphorous
runoff (q), soil moisture (moisture), precipitation (ppt), maximum Temperature (T_{max}). PERMANOVA analysis
The values marked with red are significant.

Summary CAP analysis Subsytems vs. environmental variables

Variables	p-value	CAP1 (r)	CAP2 (r)
pH	0.001	-0.743388	-0.48557
SOC	0.001	0.5641573	-0.22026
Pav	0.059	0.5685315	-0.29112
CEC	0.057	0.15248293	-0.15898
BD	0.001	-0.520975	-0.11709
Clay	0.001	-0.04241	0.230605
Sand	0.001	0.1199085	0.072694
Silt	0.07	-0.1405	-0.30992
aet	0.001	0.6470976	0.182659
q	0.001	0.569242	0.489641
mositure	0.002	0.535826	0.476092
ppt	0.128	0.6497948	0.359254
Tmax	0.01	-0.160555	0.663534

Significant level: < 0.05

Summary Enzymes abundance vs. environmental variables

Variables	p-value	CAP1 (r)	CAP2 (r)
pH	0.001	-0.6973	-0.27085
SOC	0.009	0.5183	0.20957
Pav	0.004	0.5169	0.11622
CEC	0.086	0.2813	0.03614
BD	0.001	-0.4785	-0.33867
Clay	0.003	0.1162	0.18492
Sand	0.001	-0.3984	-0.0248
Silt	0.001	0.4854	-0.12494
aet	0.03	0.6699	0.37789
q	0.002	0.7218	0.39268
mositure	0.079	0.8154	0.11556
ppt	0.014	0.7442	0.41054
Tmax	0.001	0.1773	-0.40276

Significant level: < 0.05

Summary KR-CAP

Variables	Nsap – A			Nsap- B	
	p-value	CAP1 (r)	CAP2 (r)	p-value	CAP1 (r)
pH	0.03	0.05717	0.2724	0.12	0.47
SOC	0.05	-0.1112	-0,1827	0.52	-0.64
Pav	0.20	-0,1772	-0,1745	0.19	-0.60
CEC	0.00	-0,3802	0,2268	0.91	-0.43
BD	0.03	0,0921	0,1284	0.00	0.65
Clay	0.06	0,08635	-0,3534	0.65	-0.30

Hoja1

	p-value			p-value	
Sand	0.00	0,1726	0,3906	0.03	0.40
Silt	0.00	-0,3365	-0,2576	0.47	-0.34
aet	0.00	0,01457	-0,4331	0.85	-0.59
q	0.05	-0,01404	-0,4632	0.00	-0.80
mosuture	0.07	-0,2714	-0,5149	0.16	-0.68
ppt	0.08	0,0009	-0,4779	0.41	-0.75
Tmax	0.00	-0,2399	-0,1687	0.51	0.09

	p-value	% variance	p-value	% variance
CAP model	0.001	33.7	0.05	32.6
CAP 1 axis	0.001	11.54	0.001	11.6
CAP2 axis	0.01	4.2	0.889	4.1

Significant level: < 0.05

	PhoA			PhoD	
Variables	**p-value**	**CAP1 (r)**	**CAP2 (r)**	**p-value**	**CAP1 (r)**
pH	0.07	-0.03522	-0.29193	0.001	-0.7458
SOC	0.10	-0.18357	0.11853	0.001	0.2215
Pav	0.33	0.11462	-0.35058	0.057	0.1815
CEC	0.02	0.04626	0.45434	0.001	-0.2168
BD	0.43	0.43246	-0.05363	0.001	-0.3912
Clay	0.07	-0.16675	-0.03886	0.002	0.1614
Sand	0.57	-0.34526	0.38053	0.001	-0.2812
Silt	0.46	-0.02365	0.38981	0.001	0.2735
aet	0.02	0.11362	0.45529	0.001	0.7788
q	0.58	0.15956	0.39767	0.05	0.7199
mosuture	0.58	-0.10224	0.10328	0.033	0.7656
ppt	0.46	0.14536	0.45321	0.092	0.8016
Tmax	0.06	-0.17885	-0.51421	0.001	0.313

	p-value	% variance	p-value	% variance
CAP model	0.001	13	0.001	49.8
CAP 1 axis	0.026	8.7	0.001	19.8
CAP2 axis	0.568	3.9	0.001	12.9

			BPP		
Variables	**CAP2 (r)**	**p-value**	**CAP1 (r)**	**CAP2 (r)**	**p-value**
pH	-0.08319	0.001	0.43753	0.05546	0.01
SOC	-0.36845	0.066	-0.28013	-0.4779	0.193
Pav	-0.30488	0.048	-0.33169	-0.52562	0.007
CEC	-0.25386	0.001	-0.36934	-0.61884	0.01
BD	0.17916	0.002	0.26955	0.3567	0.122
Clay	0.02404	0.022	-0.09218	0.04861	0.526
Sand	-0.05236	0.001	0.27865	0.2573	0.222
Silt	0.06312	0.002	-0.35365	-0.46807	0.302
aet	-0.11494	0.004	-0.30758	-0.36789	0.928
q	-0.45532	0.012	-0.36536	0.06208	0.002
mosuture	-0.11568	0.169	-0.62166	-0.22685	0.03
ppt	-0.30633	0.164	-0.35532	-0.1892	0.42
Tmax	0.04016	0.001	-0.28546	0.77029	0.508

	p-value	% variance	p-value

Hoja1

CAP model	0.001	47	0.03
CAP 1 axis	0.001	23	0.01
CAP2 axis	0.002	5	0.1

1
2   Hoja1
3
4   AST level 2 Subsystem annotations and Enzyme abundance and based on Kantorovich-Rubinstein dis
5   estimated (Pav), Cation Exchange Capacity (CEC), Bulk Density (BD), Clay, Sand, silt, actual evapotra
6   s with 999 permutations was performed to determine the significance between the sites/MG-RAST proje
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52   **Nsap-C**
53   **CAP2 (r)**   **p-value**
54   -0.28   0.01   0.140754   0.12030
55   -0.19   0.02   0.022451   -0.26188
56   -0.27   0.11   -0.021009   -0.27642
57   -0.07   0.16   0.015907   -0.00817
58   0.16   0.03   0.069241   0.43017
59   0.14   0.00   0.184619   -0.68596
60

Hoja1

-0.05	0.03	-0.020522	0.52716
-0.06	0.10	-0.160945	-0.12524
0.32	0.00	0.169356	-0.20919
0.27	0.02	0.003874	-0.36095
0.19	0.04	-0.16822	2 -0.45512
0.32	0.01	0.09004	9 -0.30461
0.31	0.00	-0.49931	1 -0.03969

p-value	
0.001	35.7
0.001	13.9
0.009	4.7

		PhoX
CAP2 (r)	**p-value**	**CAP1 (r)**
0.05607	0.008	0.12366
-0.09641	0.007	-0.04539
-0.14912	0.038	-0.07951
-0.17404	0.041	-0.08377
0.16044	0.074	0.02727
-0.01658	0.01	0.27152
0.16174	0.017	-0.0915
-0.22271	0.142	-0.09713
0.01391	0.001	0.27686
-0.071	0.055	0.00448
-0.2901	0.085	-0.17901
-0.0304	0.146	0.15543
-0.4091	0.001	-0.61055

p-value	**% variance**
0.001	41.4
0.001	19
0.05	4.7

Cphy	
CAP1 (r)	**CAP2 (r)**
0.56979	0.289629
-0.26451	-0.17929
-0.33896	-0.164919
-0.32083	-0.417049
0.36079	0.104345
0.18956	-0.038679
-0.11115	0.128107
0.02804	-0.159731
-0.48723	-0.474511
-0.33336	-0.375057
-0.67975	-0.40317
-0.43992	-0.45665
0.14755	0.00358

% variance

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Hoja1

49.5
10.7
7.6

Hoja1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

stance f
anspiration (aet),
ect.
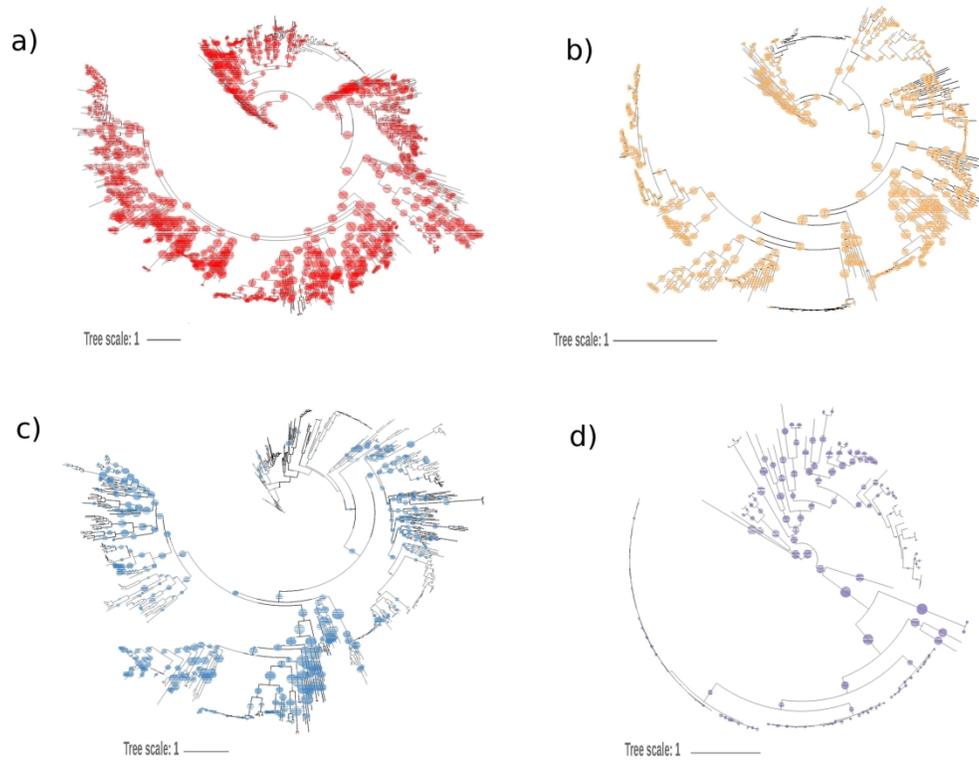
Figure 1: Phylogenetic placements of the predicted proteins of each metagenome with respect to the reference bases of each enzyme: a) PhoD, b) PhoX, c) Nsap-A and d) BPP The size of the circle representing placements is proportional to the abundance. Maximum likelihood-based phylogenetic placement of metagenome-derived protein sequences was performed with EPA-ng and tree drawn with iTOL. The circle sizes represent the number of hits per node.

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 tree drawn with iTOL. The circle sizes represent the number of hits per node.

183x178mm (300 x 300 DPI)

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); mgp9904 (violet); mgp10450 (dark blue);
mgp10523 (stone blue); mgp10541 (turquoise); mgp10956 (yellow);
mgp13011 (lilac); mgp13520 (jade); mgp13948(orange);
mpg20922 (brown); mgp89409 (brick-red); mgp91922. 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.

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); mgp9904 (violet); mgp10450 (dark blue); mgp10523 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mgp13948(orange); mpg20922 (brown); mgp89409 (brick-red); mgp91922. (light green); mgp93346 (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.

81x113mm (300 x 300 DPI)

a) alkaline phosphatases

b) Non-specific acid phosphatases
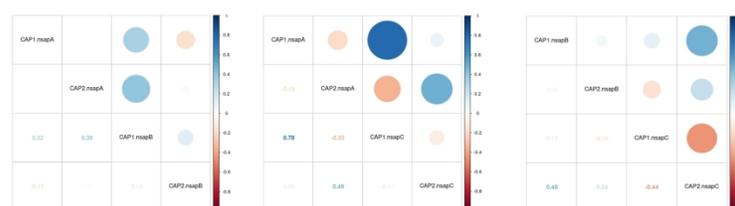
c) Phytases
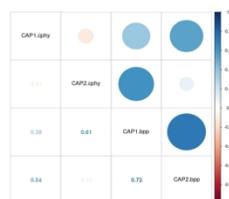
d) Most abundant

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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red the negative values. d) Correlogram of the most abundance displays the Pearson correlation coefficients. The correlation coefficients are colored according to their values; being blue the positives values and red the negative values. Correlation analysis and graphics were performed with cor R package.

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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red 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 colored according to their values; being blue the positives values and red the negative values. d) Correlogram of the most abundance displays the Pearson correlation coefficients. The correlation coefficients are colored according to their values; being blue the positives values and red the negative values. Correlation analysis and graphics were performed with cor R package.

191x271mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 (gray); mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mgp10523 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange); mpg20922 (brown); mgp89409 (brick-red); mgp91922 (light green); mgp93346 (light blue). b) The phylogeny distribution of PhoD hits along the first and second axis of the analysis (protein 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).

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 (gray); mpg8624 (mustard); mgp9904 (violet); mgp10450 (dark blue); mgp10523 (stone blue); mgp10541 (turquoise); mgp10956 (yellow); mgp13011 (lilac); mgp13520 (jade); mpg13948(orange); mpg20922 (brown); mgp89409 (brick-red); mgp91922 (light green); mgp93346 (light blue). b) The phylogeny distribution of PhoD hits along the first and second axis of the analysis (protein 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).

171x77mm (300 x 300 DPI)