Title

Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil

Journal

Nature Ecology & Evolution, 2(3)

ISSN

2397-334X

Authors

Yao, Qiuming

Li, Zhou

Song, Yang

et al.

Publication Date

2018-03-01

DOI

10.1038/s41559-017-0463-5

Peer reviewed

Community proteogenomics reveals the systemic impact of phosphorus availability on microbial functions in tropical soil

Qiuming Yao[1], Zhou Li[1], Yang Song[1], S. Joseph Wright[2], Xuan Guo[3], Susannah G. Tringe[4], Malak M. Tfaily[5], Ljiljana Paša-Tolić[5], Terry C. Hazen[1,3], Benjamin L. Turner[2], Melanie A. Mayes[1] and Chongle Pan[1,3]*

1 Oak Ridge National Laboratory, Oak Ridge, TN, USA.

2 Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama.

3 University of Tennessee, Knoxville, TN, USA.

4 Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA.

5 Pacific Northwest National Laboratory, Richland, WA, USA.

*e-mail: panc@ornl.gov

Abstract

Phosphorus is a scarce nutrient in many tropical ecosystems, yet how soil microbial communities cope with growth-limiting phosphorus deficiency at the gene and protein levels remains unknown. Here, we report a metagenomic and metaproteomic comparison of microbial communities in phosphorus-deficient and phosphorus-rich soils in a 17-year fertilization experiment in a tropical forest. The large-scale proteogenomics analyses provided extensive coverage of many microbial functions and taxa in the complex soil communities. A greater than fourfold increase in the gene abundance of 3-phytase was the strongest response of soil communities to phosphorus deficiency. Phytase catalyses the release of phosphate from phytate, the most recalcitrant phosphorus-containing compound in soil organic matter. Genes and proteins for the degradation of phosphorus-containing nucleic acids and phospholipids, as well as the decomposition of labile carbon and nitrogen, were also enhanced in the phosphorus-deficient soils. In contrast, microbial communities in the phosphorus-rich soils showed increased gene abundances for the degradation of recalcitrant aromatic compounds, transformation of nitrogenous compounds and assimilation of sulfur. Overall, these results demonstrate the adaptive allocation of genes and proteins in soil microbial communities in response to shifting nutrient constraints.

Introduction

The majority of soil phosphorus is sequestered in recalcitrant minerals and organic compounds that are not directly available for biological assimilation. A number of strategies have been developed by soil microorganisms and plants to access limited phosphorus reservoirs in soil, including secretion of weak acids to solubilize phosphorus from the soil matrix and synthesis of phosphatase enzymes to release phosphorus from organic compounds[1,2]. Biological phosphorus acquisition from these refractory substrates is slow and requires significant investment of metabolic resources. Because

phosphorus availability limits primary productivity and carbon cycling in tropical terrestrial ecosystems, detailed understanding of phosphorus acquisition by soil microbial communities is critical for ecological and climate modelling.

The ecological effects of different levels of phosphorus availability have been investigated in a tropical forest located on the Gigante Peninsula in the Barro Colorado Nature Monument, Republic of Panama. Phosphorus is a key limiting nutrient in this ecosystem. A long-term fertilization experiment, which has been ongoing since 1998, has increased the concentration of soluble phosphorus from $<1\,mg\,kg^{-1}$ in the control plots to $>30\,mg\,kg^{-1}$ in the phosphorus-fertilized plots[3]. The soluble phosphorus concentrations in the control plots are among the lowest of soils in central Panama and are comparable to extremely infertile soils in lowland tropical forests worldwide, including the entire Amazon Basin[4]. In contrast, the $>30\,mg\,kg^{-1}$ of soluble phosphorus in the phosphorus-fertilized plots exceeds the highest natural soil phosphorus concentrations measured in lowland forests of Panama. Thus, the control and phosphorus-fertilized plots in the Gigante fertilization experiment span the regional range in available phosphorus and are comparable to the range in available phosphorus in lowland tropical forests worldwide. The observed effects of phosphorus fertilization on plants included greater fine-litter production[5,6], higher tissue phosphorus concentrations[6,7] and lower phosphorus resorption proficiency[8]. The more abundant phosphorus also increased microbial biomass in soil and decreased the enzyme activities of phosphatase and *N*-acetyl β-glucosaminidase[3].

The ecological impacts of phosphorus availability on soil microbial communities have also been investigated using long-term phosphorus fertilization in other natural terrestrial ecosystems, including a tropical forest in southern China[9], a grassland in the Netherlands[10], and arctic tundra in northern Sweden[11,12] and Alaska[13,14]. In general, these studies found that phosphorus fertilization increased soil microbial biomass and nutrient content, altered soil microbial diversity, and suppressed phosphatase activities. However, it remains unknown how phosphorus deficiency enhances microbial acquisition of phosphorus from organic compounds and which taxa in soil communities drive such enhancement. Acquisition of organic phosphorus is linked with decomposition of soil organic matter that also contains organic nitrogen and carbon, but there is little information about the interactions between metabolic processes for phosphorus acquisition, nitrogen uptake and carbon degradation.

Integrated -omics studies have revealed key soil microbial activities in diverse ecosystems, including methanotrophy in Arctic peat soils[15], methanogenesis in permafrost soils[16] and methylotrophy in grassland soils[17]. Here, we report a proteogenomics comparison of soil microbial communities with and without long-term phosphorus fertilization. Our study used deeper sequencing of soil metagenomes (an average of 157 gigabase pairs (Gbp) of raw sequencing data per soil sample yielding 2.3 Gbp of assembly) and

generated more protein identifications from soil metaproteomes (an average of 7,114 protein identifications per soil sample) than previous studies[15,16,17]. In combination, community proteogenomic analyses provided a comprehensive survey of the adaptation of soil communities to different phosphorus levels through the gain and loss of metabolic capabilities, shifts in community structure and regulation of enzyme abundances.

Results

Changes of gene abundances for key nutrient cycling functions by phosphorus availability

Soil samples were collected from two phosphorus-deficient control plots (plots 6 and 36) and two phosphorus-rich plots after 17 years of phosphorus fertilization (plots 1 and 30). Supplementary Data 1 summarizes the overall results of the proteogenomic measurements. Read-based analyses of metagenomic data were used to quantify the gene abundances of enzyme commission (EC) numbers and Gene Ontology (GO) terms in the soil communities (Supplementary Data 2). Statistical comparison between the phosphorus-deficient and phosphorus-rich soils identified EC numbers and GO terms with significantly different gene abundances (that is, q values <0.05 and abundance changes ≥ 1.2-fold). There were 74 EC numbers and 135 GO terms with increased gene abundances in the phosphorus-rich soils and 45 EC numbers and 81 GO terms with increased gene abundances in the phosphorus-deficient soils.

Many genes for phosphorus acquisition were more abundant in the phosphorus-deficient soils than the phosphorus-rich soils (Fig. 1a and Table 1). The gene abundance of 3-phytase for extraction of phosphates from phytate was 4.5 times higher in the phosphorus-deficient soils than the phosphorus-rich soils. The phosphorus-deficient soils also had more than twofold higher gene abundances than the phosphorus-rich soils for sphingomyelin phosphodiesterase, phospholipase A1 and phospholipase A2. The gene abundances for exoribonucleases were 1.6-fold higher in the phosphorus-deficient soils than the phosphorus-rich soils. These findings indicated that microbial communities in the phosphorus-deficient soils enhanced their genetic capabilities to recycle phosphorus from phytate, phospholipids and nucleic acids in soil organic matter. In particular, the gene abundance increase for 3-phytase was the highest among all enzymes enhanced by phosphorus deficiency, suggesting that phytate is a key phosphorus source in the Gigante soils.

Fig. 1: Impact of phosphorus availability on the phosphorus, carbon, nitrogen and sulfur cycles of the Gigante soil communities. a, Phosphorus acquisition from organic phosphorus substrates ordered by recalcitrance. b, Degradation of carbon substrates ordered by recalcitrance. c, Transformation of nitrogenous compounds ordered by nitrogen reduction states. d, Transformation of sulfur substrates ordered by sulfur reduction states. EC numbers associated with each transformation process are shown with red arrows when more abundant in phosphorus-deficient soils, blue arrows when more abundant in phosphorus-rich soils and black arrows when present in both soils with no significant abundance change. Microbial communities in the phosphorus-deficient soils had higher gene abundances for enzymes (Table 1) to acquire phosphorus from phytate, nucleic acids and phospholipids, decompose hemicellulose and pectin, and degrade proteins. Microbial communities in the phosphorus-rich soils had higher gene abundances for enzymes (Table 1) to catabolize aromatic compounds, obtain ammonium from nitrophenol and ethanolamine, and uptake sulfur from taurine.

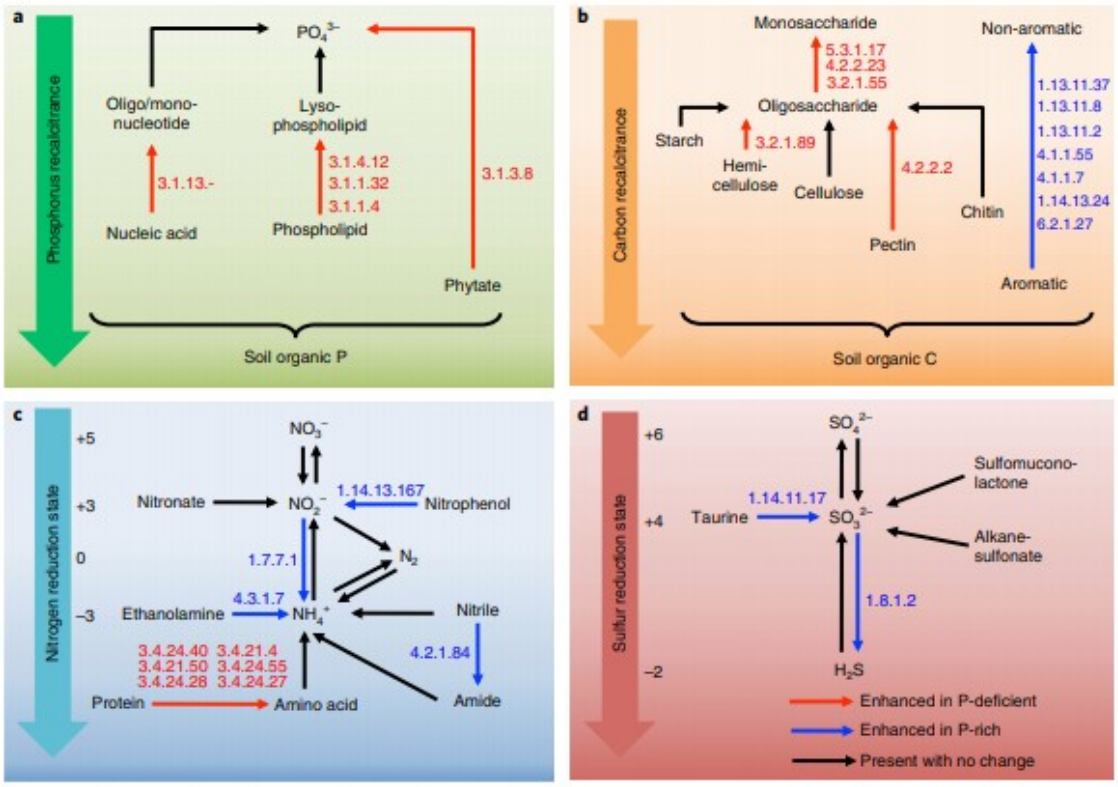


Table 1 | Gene abundance changes for key enzymes involved in nutrient cycling

Nutrient	Enzyme name[a]	EC number	q value	Folds
Enhanced in phosphorus-deficient soils				
Phosphorus cycling	3-phytase	3.1.3.8	3.8×10^{-4}	4.5
	Sphingomyelin phosphodiesterase	3.1.4.12	1.5×10^{-2}	3.2
	Phospholipase A1	3.1.1.32	2.7×10^{-2}	2.2
	Phospholipase A2	3.1.1.4	3.2×10^{-2}	2.2
	Exoribonucleases	3.1.13.-	3.3×10^{-2}	1.5
Carbon cycling	Pectate lyase	4.2.2.2	3.3×10^{-2}	1.9
	Arabinogalactan endo-β-1,4-galactanase	3.2.1.89	2.3×10^{-2}	2.0
	Rhamnogalacturonan endolyase	4.2.2.23	6.2×10^{-3}	2.5
	5-dehydro-4-deoxy-D-glucuronate isomerase	5.3.1.17	1.2×10^{-2}	1.8
	α-L-arabinofuranosidase	3.2.1.55	7.1×10^{-3}	1.2
Nitrogen cycling	Pitrilysin	3.4.24.55	2.0×10^{-3}	2.3
	Lysyl endopeptidase	3.4.21.50	6.5×10^{-4}	2.1
	Bacillolysin	3.4.24.28	1.7×10^{-4}	2.0
	Thermolysin	3.4.24.27	3.7×10^{-3}	1.9
	Serralysin	3.4.24.40	3.9×10^{-4}	1.8
	Trypsin	3.4.21.4	2.3×10^{-3}	1.4
Enhanced in phosphorus-rich soils				
Carbon cycling	4-hydroxybenzoate-CoA ligase	6.2.1.27	2.6×10^{-2}	1.3
	3-hydroxybenzoate 6-monooxygenase	1.14.13.24	9.5×10^{-4}	1.3
	Benzoylformate decarboxylase	4.1.1.7	6.2×10^{-3}	1.2
	4,5-dihydroxyphthalate decarboxylase	4.1.1.55	3.8×10^{-4}	1.3
	Catechol 2,3-dioxygenase	1.13.11.2	9.5×10^{-4}	1.2
	Protocatechuate 4,5-dioxygenase	1.13.11.8	4.1×10^{-2}	1.2
	Hydroxyquinol 1,2-dioxygenase	1.13.11.37	1.9×10^{-2}	1.2
Nitrogen cycling	4-nitrophenol 4-monooxygenase	1.14.13.167	8.7×10^{-6}	1.3
	Ferredoxin-nitrite reductase	1.7.7.1	4.8×10^{-3}	1.2
	Ethanolamine ammonia-lyase	4.3.1.7	5.0×10^{-2}	1.2
	Nitrile hydratase	4.2.1.84	2.7×10^{-3}	1.2
Sulfur cycling	Taurine dioxygenase	1.14.11.17	2.2×10^{-6}	1.3
	Assimilatory sulfite reductase	1.8.1.2	4.3×10^{-4}	1.3

[a]The enzymes are also shown in Fig. 1 by their EC numbers.

Phosphorus deficiency altered the gene abundances of many enzymes for the degradation of carbon substrates[18] (Fig. 1b and Table 1). The phosphorus-deficient soils had higher gene abundances than the phosphorus-rich soils for pectate lyase and arabinogalactan endo-β-1,4-galactanase. For the further degradation of oligosaccharides to monosaccharides, the phosphorus-deficient soils had higher gene abundances for rhamnogalacturonan endolyase, 5-dehydro-4-deoxy-D-glucuronate isomerase, and α-L-arabinofuranosidase. In contrast, the phosphorus-rich soil communities had higher gene abundances than the phosphorus-deficient soil communities for many enzymes involved in the degradation of aromatic compounds, including 4-hydroxybenzoate-CoA ligase, 3-hydroxybenzoate 6-monooxygenase, benzoylformate decarboxylase and 4,5-dihydroxyphthalate decarboxylase. Oxygenases capable of opening aromatic rings also had higher gene abundances in the phosphorus-rich soils; for example, catechol 2,3-dioxygenase, protocatechuate 4,5-dioxygenase and hydroxyquinol 1,2-dioxygenase. This suggests that the microbial

communities degraded relatively higher amounts of recalcitrant aromatic carbon in the phosphorus-rich soils and relatively higher amounts of labile polysaccharide carbon in the phosphorus-deficient soils.

Phosphorus fertilization also altered the gene abundances of enzymes involved in nitrogen transformations (Fig. 1c and Table 1). The phosphorus-deficient soils had higher gene abundances than the phosphorus-rich soils for many proteases, including pitrilysin, lysyl endopeptidase, bacillolysin, thermolysin, serralysin and trypsin. In contrast, the phosphorus-rich soils had higher gene abundances than the phosphorus-deficient soils for nitrile hydratase, ethanolamine ammonia-lyase and nitrophenol monooxygenase. Nitrite reductase for assimilatory nitrite reduction to ammonia was also significantly more abundant in the phosphorus-rich soils than the phosphorus-deficient soils. Thus, the phosphorus-rich communities may have a higher demand to obtain nitrogen from a variety of nitrogenous substrates than the phosphorus-deficient communities. Conversely, the phosphorus-deficient communities may rely more on amino acids in decaying proteins for nitrogen.

For sulfur cycling (Fig. 1d and Table 1), the phosphorus-rich soil communities had significantly higher gene abundances than the phosphorus-deficient soil communities for taurine dioxygenase and assimilatory sulfite reductase. This suggested a higher level of sulfur assimilation by the phosphorus-rich soil communities than the phosphorus-deficient soil communities.

Assembly and binning of high-quality genomes from the metagenomics data

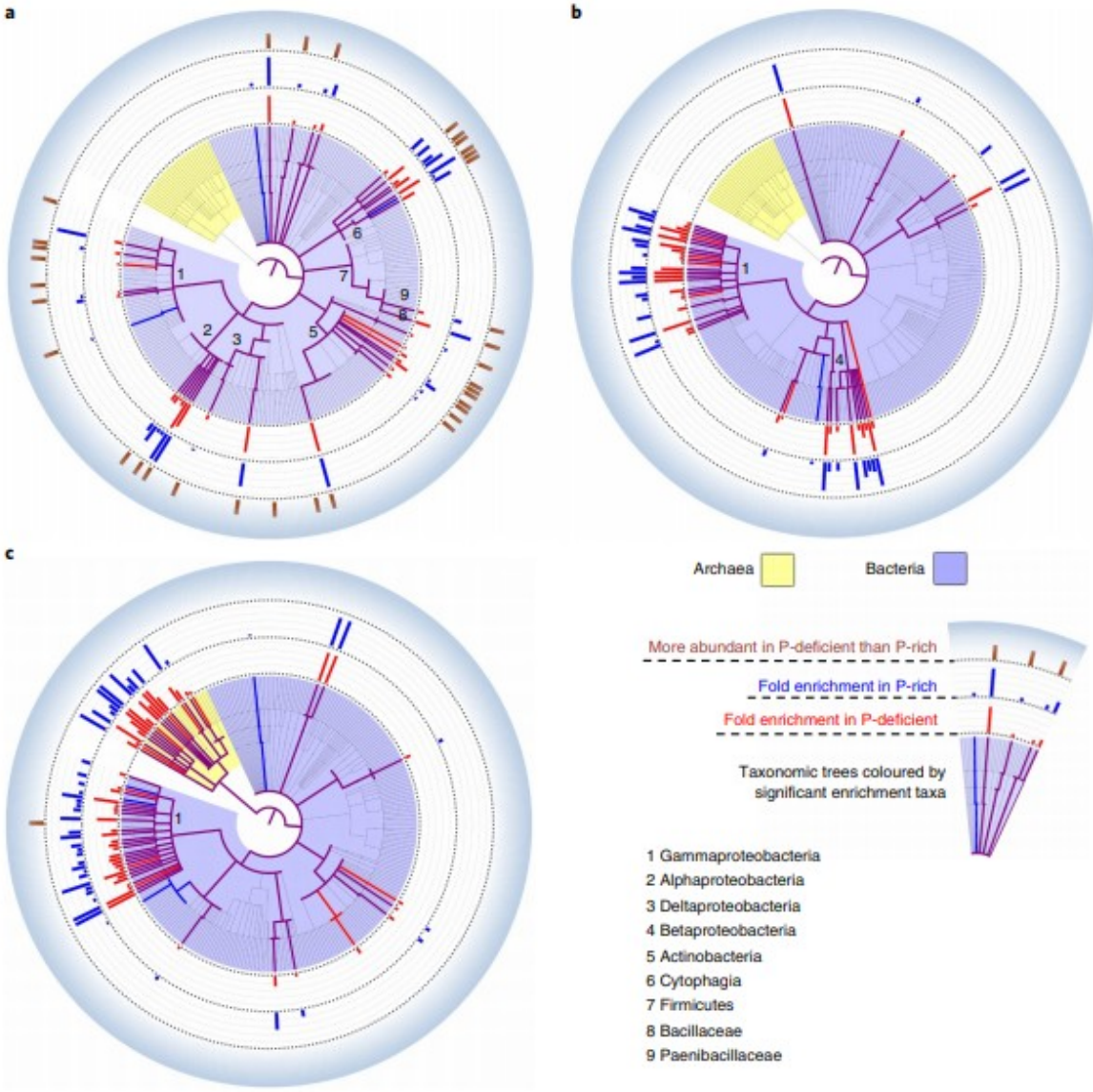
Assembly and binning generated 12 new genomes with more than 90% completeness and less than 10% contamination (Supplementary Data 3). These near-complete genomes are numbered from G1 to G12 in Supplementary Figs. 1 and 2 and represent bacteria in a novel phylum (G1), the Acidobacteria phylum (G9), the Cyanobacteria phylum (G10), two in the Planctomycetacia class (G11 and G12), the Actinobacteria class (G7), the Deltaproteobacteria class (G4), four in the Myxococcales order (G2, G3, G5 and G6) and one in the Acidobacteriales order (G8). There were 69 common metabolic pathways identified in these novel genomes (Supplementary Data 3). Many of these common pathways were involved in the metabolism of fatty acids, amino acids and nucleic acids. Two genomes in the Myxococcales order (G2 and G3) contained genes for phytases and phospholipases involved in phosphorus acquisition. Assimilatory sulfate reduction was found in two genomes of the Deltaproteobacteria class (G4 and G6). Dissimilatory nitrate reduction was identified in the genome of the Cyanobacteria phylum (G10).

Taxonomic and phylogenetic analyses of phytases in the Gigante soil communities

The metagenomic data were used to infer the taxonomic distribution of the key phosphorus acquisition enzymes enhanced by phosphorus deficiency,

including 3-phytase (Fig. 2a), phospholipases (Fig. 2b) and exoribonuclease phosphomonoesterase (Fig. 2c). A previous study has found Bacillaceae and Paenibacillaceae to have phytase genes and phytate degradation activities[19]. Here, we also found 3-phytase genes to be enriched in Bacillaceae and Paenibacillaceae by approximately ten- and fourfold, respectively. The consistency of our findings with the previous study supported the validity of our taxonomy enrichment analysis based on metagenomics. In addition to Bacillaceae and Paenibacillaceae, the 3-phytase genes were enriched in 31 other families in the Gigante soil communities, among which Promicromonosporaceae, Desulfuromonadaceae, Kordiimonadaceae, Parvularculaceae, Chlorobiaceae and Cytophagaceae had >15-fold enrichments. Most of the families with enriched 3-phytase genes also had higher gene abundances of 3-phytase in the phosphorus-deficient soils than the phosphorus-rich soils. This suggests that instead of selecting for a few specific phytate-degrading families, phosphorus deficiency increased the abundances of most organisms capable of phytate degradation. While families enriched in 3-phytase genes were dispersed across the taxonomic tree, the phospholipase genes were enriched in the Gammaproteobacteria and Betaproteobacteria classes (Fig. 2b), and the exoribonuclease phosphomonoesterase genes were primarily enriched in the Gammaproteobacteria class and the Archaea domain (Fig. 2c).

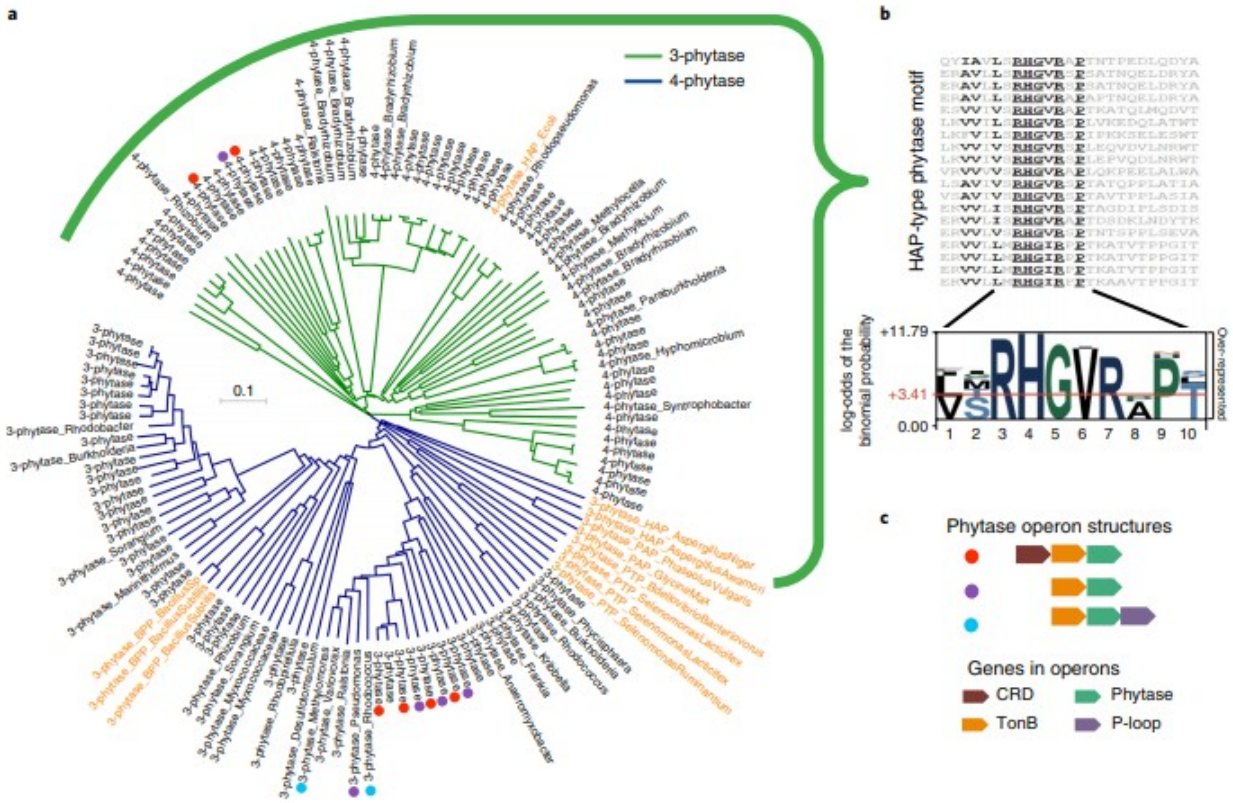
Fig. 2: Taxonomic distribution of key phosphorus-acquisition enzymes. a, 3-phytase for phytate. b, Phospholipases A1 and A2 for phospholipids. c, Exoribonuclease phosphomonoesterase for nucleic acids. The taxonomic tree includes 265 families in the Bacteria (blue background) and Archaea (yellow background) domains. The branches are highlighted in red for enriched families in the phosphorus-deficient communities, blue for enriched families in the phosphorus-rich communities and purple for enriched families in both communities. The fold enrichment for a family is shown in the ring of red bars for the phosphorus-deficient communities and the ring of blue bars for the phosphorus-rich communities. The families with significantly different gene abundances between the phosphorus-deficient and phosphorus-rich communities are marked in the outermost ring of brown bars. Key branches of the taxonomic tree are labelled with numbers and identified by their names.



The four Gigante soil metagenomes encoded 57 genes for 3-phytase and 57 genes for 4-phytase. The 4-phytase genes were clearly separated from the 3-phytase genes in the phylogenetic tree (Fig. 3a). Most 4-phytase genes contained a conserved sequence motif of RHG-R-P (Fig. 3b), which is a hallmark of histidine acid phosphatase[20]. Operons for 40 of these phytase genes were identified from large scaffolds. The common operon structures shown in Fig. 3c indicate that 13 phytase genes were located downstream of a bacterial TonB receptor gene. This is consistent with the findings of a previous study on phytases[21], which hypothesized the adjacent TonB receptor gene to be involved in phytate transport.

Fig. 3: Classification and evolution of phytase genes. a, Phylogenetic tree of 3-phytase (blue branches) and 4-phytase (green branches). Genes in black are new phytases identified in the four Gigante soil communities and genes in orange are reference phytases from Swiss-Prot. HAP, histidine acid phosphatases; BPP, β-propeller phytases; PTP, protein tyrosine

phosphatase-like phytases. b, Conserved sequence motif for histidine acid phosphatases. Most new 4-phytases are histidine acid phosphatases. c, Operon structures of the phytase genes. Common genes in the phytase-containing operons include CRD (carboxypeptidase regulatory domain), TonB (bacterial TonB-dependent receptor, β-barrel) and P-loop (phosphorus loop containing nucleoside triphosphate hydrolase). These genes are typically arranged in three operon structures, which are highlighted by red, purple and blue dots in a.

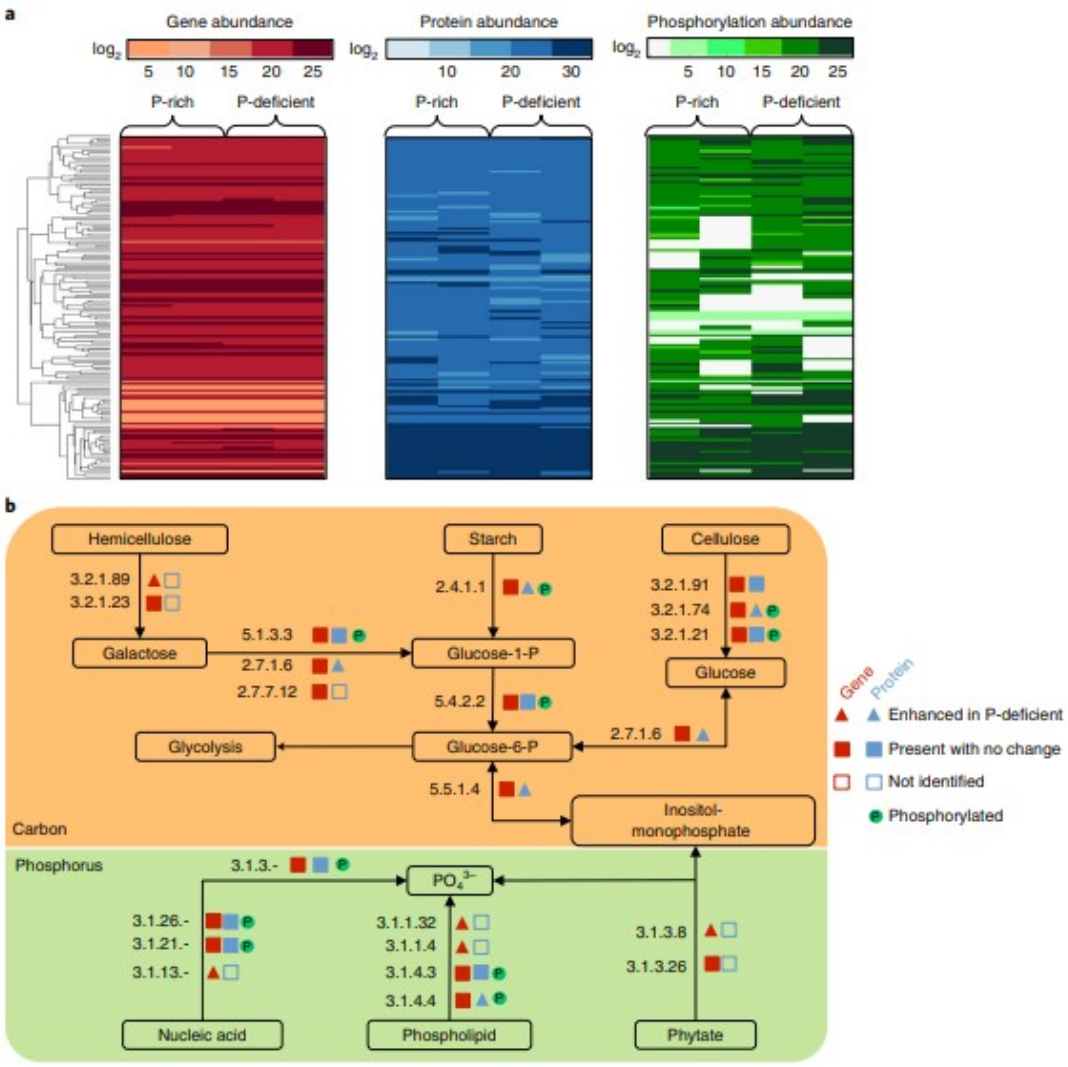


Metaproteomics of the Gigante soil communities

Metaproteomics results were summarized in Fig. 4a and Supplementary Data 1. Statistical comparison between the phosphorus-deficient and phosphorus-rich metaproteomes found 23 EC numbers and 14 GO terms with higher protein abundances in the phosphorus-rich soils, and 11 EC numbers and 16 GO terms with higher protein abundances in the phosphorus-deficient soils at q values less than 0.05 (Supplementary Data 4). Starch phosphorylase and glucan 1,4-β-glucosidase had higher protein abundances in the phosphorus-deficient soils than in the phosphorus-rich soils (Fig. 4b). Two additional cellulases—cellulose 1,4-β-cellobiosidase and β-glucosidase—were identified by metaproteomics without significant change. Furthermore, metagenomics identified endo-β-1,4-galactanase and β-galactosidase, which act together to degrade arabinogalactan in hemicellulose to galactose. Galactose can be converted to glucose 1-phosphate through the Leloir pathway consisting of galactose mutarotase, galactokinase and D-galactose-1-phosphate uridylyltransferase. Three of these enzymes were detected by metagenomics and two by metaproteomics. Galactokinase had a higher protein abundance

in the phosphorus-deficient soils than the phosphorus-rich soils. Taken together, the proteogenomics results (Fig. 4b) indicated that the phosphorus-deficient communities had higher degradation capabilities for polysaccharides, including starch, cellulose and hemicellulose, than the phosphorus-rich communities.

Fig. 4: Proteogenomics comparison of phosphorus-deficient and phosphorus-rich soils. a, Heatmaps of gene abundances (red), protein abundances (blue) and phosphorylation abundances (green) of identified enzymes in the four soils. b, Regulation of enzymes involved in carbon degradation (orange background) and phosphorus acquisition (green background). The enzymes are marked with triangles for increased abundances by phosphorus deficiency, filled squares for identified abundances with no significant change and open squares for no identified abundances. The symbols are in red for gene abundances and blue for protein abundances. Enzymes with phosphorylation detected are marked with green circles. Starch phosphorylase (EC 2.4.1.1, q value = 1.0×10^{-2}), glucan 1,4-β-glucosidase (EC 3.2.1.74, q value = 4.5×10^{-2}), galactokinase (EC 2.7.1.6, q value = 4.0×10^{-2}), inositol-3-phosphate synthase (EC 5.5.1.4, q value = 4.9×10^{-2}) and phospholipase D (EC 3.1.4.4, q value = 1.0×10^{-2}) have enhanced protein abundances in the phosphorus-deficient soils for carbon degradation and phosphorus acquisition.



Inositol-monophosphate is the degradation product of phytate after five of its six phosphate groups are released by phytases. Inositol-monophosphate can be reversibly converted into glucose-6-phosphate for glycolysis by inositol-3-phosphate synthase. The protein abundance of inositol-3-phosphate synthase significantly increased in the phosphorus-deficient soils compared

with the phosphorus-rich soils. The GO term for the inositol catabolic process (GO:0019310, q value $= 4.2 \times 10^{-2}$) also had a significantly higher protein abundance in phosphorus-deficient soils. This indicated higher catabolism rates of the carbon backbone of phytate in the phosphorus-deficient communities.

Metaproteomics identified five phospholipases and nucleases involved in phosphorus acquisition, including phosphoric monoester hydrolases, endoribonucleases producing 5′-phosphomonoesters, endodeoxyribonucleases producing 5′-phosphomonoesters, phospholipase C and phospholipase D. Phospholipase D had a higher protein abundance in the phosphorus-deficient soil than the phosphorus-rich soil. All five of these enzymes carried phosphorylation. More generally, metaproteomics found a higher protein abundance for the cellular response to phosphate starvation (GO:0016036, q value $= 8.2 \times 10^{-3}$) in the phosphorus-deficient soils than the phosphorus-rich soils.

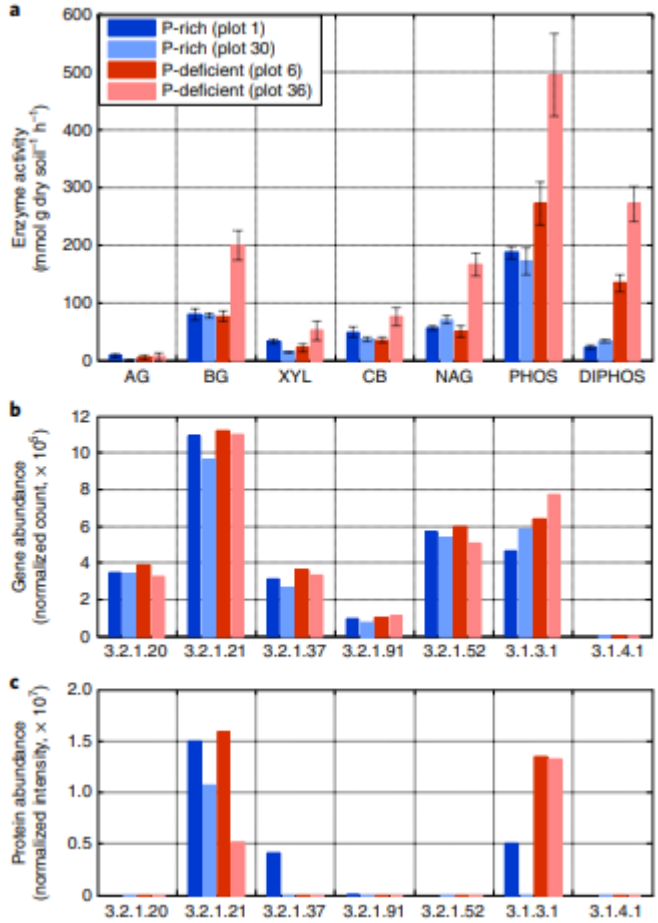
Differential analysis of the metaproteomic data based on GO terms revealed many enhanced transporting activities related to phosphorus acquisition in the phosphorus-deficient soils. Phosphate transport (GO:0035435, q value $= 3.4 \times 10^{-2}$) and phosphate ion binding (GO:0042301, q value $= 4.7 \times 10^{-2}$) had significantly higher protein abundances in the phosphorus-deficient soils than the phosphorus-rich soils. The protein abundances for phosphonate transporter activity (GO:0042917, q value $= 3.9 \times 10^{-2}$) and lipid transporting (GO:0034040, q value $= 4.7 \times 10^{-2}$) also increased in the phosphorus-deficient soils.

Soil property analyses and enzyme assays

As shown previously[3, 22,23], phosphorus addition did not significantly alter the soil pH, effective texture, total carbon and nitrogen or dissolved carbon and nitrogen, but it did increase the microbial biomass carbon and nitrogen (Supplementary Data 5). There was no clear difference between the phosphorus-rich and phosphorus-efficient soils in the fractions of different biochemical classes of compounds in soil organic matter measured by Fourier transform ion cyclotron resonance mass spectrometry (MS) (Supplementary Data 6).

Enzyme assays were used to measure the enzymatic activities of α-glucosidase, β-glucosidase, β-D-cellobiosidase, *N*-acetyl β-glucosaminidase, β-xylosidase, phosphomonoesterase and phosphodiesterase in fresh samples of the four soil communities (Fig. 5a). The phosphorus-deficient soils had significantly higher levels of phosphomonoesterase and phosphodiesterase activities than the phosphorus-rich soils, but no significant difference for the other enzyme activities. The gene abundances and the protein abundances of the representative EC numbers corresponding to these enzymatic activities[3] are shown in Fig. 5b,c, respectively. The enhancement of phosphomonoesterase in the phosphorus-deficient soils was consistently detected by metagenomics, metaproteomics and enzyme assays.

Fig. 5: Comparison of enzyme activities. a, Measurement of enzyme activities in the phosphorus-deficient soils (red) and phosphorus-rich soils (blue), including α-glucosidase (AG; EC 3.2.1.20), β-glucosidase (BG; EC 3.2.1.21), β-xylosidase (XYL; EC 3.2.1.37), β-D-cellobiosidase (CB; EC 3.2.1.91), *N*-acetyl β-glucosaminidase (NAG; EC 3.2.1.52), phosphomonoesterase (PHOS; EC 3.1.3.1) and phosphodiesterase (DIPHOS; EC 3.1.4.1). The error bars were estimated from the triplicate enzyme assays on each soil. b,c, Gene (b) and protein abundances (c) of the above representative EC numbers for these enzyme activities.



Discussion

Phosphorus deficiency enhanced the gene abundances of specific phosphatases for phytate, nucleic acids and phospholipids (Fig. 1a). In particular, phytase had a much higher increase in gene abundance than phosphatases for nucleic acids and phospholipids. This supports the hypothesis that phosphorus deficiency drives soil microbial communities to extract phosphorus from more recalcitrant substrates[24]. This result highlights the importance of microbial degradation of phytate—relative to other organic phosphorus compounds—in phosphorus-deficient soils, and provides a mechanism to explain low concentrations of phytate in lowland tropical forest soils[25].

The taxonomic distribution of phytase genes was similar between the phosphorus-rich and phosphorus-deficient communities (Fig. 2). This

suggests that 17 years of phosphorus fertilization did not induce significant loss of phytase genes in the genomes of organisms originally capable of phytate degradation. However, the abundances of these organisms were significantly reduced by the phosphorus fertilization. The taxonomic distribution of phytase genes was distinct from those genes for the degradation of nucleic acids and phospholipids (Fig. 2). This supports the extrapolation of the resource partitioning hypothesis of soil phosphorus among different plant groups[26,27] to different taxa of soil microorganisms. The organisms capable of degrading nucleic acids or phospholipids did not have significantly higher abundances in phosphorus-deficient soils. This suggests that phosphorus deficiency provided substantial competitive advantages to the microorganisms capable of phytate degradation, but not to the microorganisms involved in the degradation of nucleic acids and phospholipids.

The demand for phytate degradation in this tropical ecosystem created a large diversity of phytase genes (Fig. 3). The >100 phytase genes discovered here represent a substantial addition to the currently known microbial phytase genes. Analysis of the operon structures of these phytase genes indicated their association with a TonB gene, which was hypothesized to play a role in phytate transport in a previous study[21]. Thus, our proteogenomics analyses of the phosphorus-deficient and phosphorus-rich soil communities covered the entire process of phytate degradation: first the transportation of phytate by a TonB gene, next the higher gene abundance of phytase releasing phosphates from phytate and finally the higher protein abundance of the enzymes catabolizing the carbon backbone of phytate.

Microbial communities require carbon, nitrogen and sulfur along with phosphorus in relatively fixed ratios[28], but nutrient resources are supplied in variable proportions in natural environments. Such elemental imbalances drive many ecological phenomena, as surmised by the theory of ecological stoichiometry[29,30]. Here, complementary enzyme assays, metagenomics and metaproteomics were used to provide a survey of microbial acquisition of carbon, nitrogen, phosphorus and sulfur from a variety of substrates. Our results highlighted the importance of microbial degradation of recalcitrant substrates to acquire limiting elements. In phosphorus-deficient soils, the communities had higher gene abundances for the degradation of phytate for phosphorus. In phosphorus-rich soils, the microbial communities had higher gene abundances for the degradation of aromatic compounds, transformation of various nitrogenous compounds for nitrogen and degradation of taurine for sulfur. So far, there has been no significant change in soil organic carbon and nitrogen concentrations in the phosphorus-fertilized plots[23], despite a greater rate of decomposition[31]. This suggests that the additional capacity of the microbial community to degrade recalcitrant organic matter has been offset by the greater input (25% increase) of litter fall in the phosphorus-fertilized plots[5]. Overall, our results suggest that the selective degradation of recalcitrant substrates may be an important

ecological stoichiometry mechanism for microbial communities to manage elemental imbalances in soils.

Adaptive allocation of genes and proteins for resource acquisition by microbial communities suggests optimal foraging. Optimal foraging theory[32,33] predicts that an efficient microbial community should adjust its foraging effort for different nutrients to maximize the acquisition of limiting nutrients and avoid excessive acquisition of non-limiting nutrients. When a nutrient can be acquired from multiple substrates, an efficient community should extract from the least recalcitrant substrates to minimize foraging effort[34]. The foraging effort by a microbial community can be represented by the gene and protein abundances of enzymes for nutrient acquisition. Here, we found that microbial communities invested more into genes and proteins for a variety of phosphorus acquisition enzymes in phosphorus-deficient soils to maximize the foraging effort for phosphorus. The addition of soluble phosphorus alleviated phosphorus limitation, which probably enhanced the demand for other nutrients and drove the community to switch investment from the acquisition of phosphorus towards the acquisition of carbon, nitrogen and sulfur. Previously, optimal foraging theory has been used to describe plants' adaptive allocation of photosynthate or biomass for resource uptake in response to different resource limitations[35]. Here, our results suggest that optimal foraging can be generalized to the allocation of genes and proteins in microbial communities for nutrient acquisition. The allocation plasticity stems from the compositional variation of a community and the protein expression regulation of individual microorganisms.

In conclusion, our proteogenomics results provide systems biology insights into the adaptation of soil microbial communities to different levels of phosphorus availability in a humid tropical forest environment. Phosphorus deficiency significantly enhanced the genetic capabilities of microbial communities to extract phosphorus from phytate and, to a lesser extent, from nucleic acids and phospholipids. Long-term phosphorus fertilization altered the allocation of genes and proteins by microbial communities to acquire carbon, nitrogen and sulfur from a variety of substrates. The results suggest that the selective degradation of recalcitrant substrates, including phytate in phosphorus-deficient soils and aromatic compounds in phosphorus-rich soils, is an important means for microbial communities to balance their elemental requirements. The adaptive allocation of genes and proteins for acquisition of these nutrients in different soils can be explained as an optimal foraging strategy by which microbial communities maintain efficient growth under resource limitation.

Methods

Study site and soil sampling

The Gigante fertilization experiment began in 1998 on the Gigante Peninsula in the Smithsonian Tropical Research Institute, Republic of Panama, as described previously in detail[3,5]. The control plots did not receive any

fertilization and the phosphorus-fertilized plots received 50 kg phosphorus ha^{-1} $year^{-1}$ as triple super phosphate applied in four equal doses each year. In this study, the phosphorus-deficient soils were collected from two control plots (plots 6 and 36) and the phosphorus-rich soils were collected from two phosphorus-fertilized plots (plots 1 and 30) in December 2014. After litter was removed, surface soil was excavated to a depth of 10 cm from three replicate sites located ~10 m apart at the apices of a triangle in the centre of each of the four plots. Each experimental plot measured 40 × 40 m (ref. 5). The three replicate soils from each plot were homogenized and combined with equal weight into a composite soil sample for metagenomic sequencing and metaproteomic analysis. The composite soil samples were shipped in dry ice and preserved in −80 °C before DNA and protein extraction.

Metagenomic sequencing

High-molecular-weight DNA was extracted in triplicate from each of the composite soil samples using a protocol described previously[36]. Briefly, each extraction mixed 300 mg of soil with 100 mg of sand and 0.1 ml of solution A (4 M guanidine thiocyanate, 25 mM Tris (pH 7.2), 10 mM EDTA, 10 µl ml^{-1} 2-mercaptoethanol and 0.1% sarkosyl). The samples were ground to fine powder in liquid nitrogen with a mortar and pestle. The samples were then mixed with 1.4 ml of solution B (6.4 M urea, 1 M NaCl, 80 mM Tris (pH 7.0) and 40 mM EDTA) and centrifuged at 10,000 *g* for 1 min at room temperature. The supernatant was mixed with 200 µl of pre-heated solution C (2.5% cetyltrimethylammonium bromide and 2.5% sarkosyl). The pellet was re-suspended in 250 µl of solution D (500 mM Na_2HPO_4 and 500 mM NaH_2PO_4 (pH 7.2)) and centrifuged at 10,000 *g* for 1 min. The supernatant was combined with the previous supernatant. This desorption step using solution D was repeated once more with the pellet. The combined supernatant was mixed with 1.6 ml of chloroform and centrifuged at 7,500 *g* for 20 min. DNA in the aqueous phase was precipitated with 0.6 volume of isopropanol and centrifuged at 16,000 *g* for 20 min. The extracted DNA was purified using a PowerClean Pro DNA Clean-Up Kit (Mo Bio) and was further concentrated using the DNA Clean & Concentrator Kit (Zymo Research). The four obtained high-molecular-weight DNA samples were sequenced using a total of 7 lanes of Illumina HiSeq 2× 250 base pair (bp), including 2 lanes for each metagenome from plots 1, 30 and 36, and 1 lane for the metagenome from plot 6. In addition, low-molecular-weight DNA was also extracted from each composite soil sample using the PowerMax Soil DNA Isolation Kit (Mo Bio) following the manufacturer's protocol. The obtained low-molecular-weight DNA samples from the four plots were multiplexed together and sequenced by a single lane of Illumina HiSeq 2× 150 bp.

Read-based data analyses

The metagenomic data from the 2 × 250 bp sequencing of high-molecular-weight DNA was pre-processed using BBTools (http://jgi.doe.gov/data-and-tools/bbtools/) and error-corrected by read alignment using Omega version

2.0 (ref. 37) at a mismatch rate of <3%[38]. Overlapping paired-end reads were merged into consensus insert sequences using the read overlapping algorithm in Omega version 2.0 (ref. 38). Reads were searched against a reference protein database on the Titan supercomputer using DIAMOND in the blastx mode[39] with an e value threshold of 0.001. The reference protein database included proteins from representative genomes for all 4,055 prokaryotic species in the National Center for Biotechnology Information complete genome repository in November 2015. If a species had multiple genomes, the genome with the largest number of proteins was selected as the representative genome for that species. The 4,055 species belonged to 265 bacterial and archaeal families shown in the taxonomic tree of Fig. 2. The genomes were re-annotated consistently by UniFam, as described previously[40]. The gene abundance of an EC number or a GO term was quantified in a metagenome based on the base-pair counts of reads mapped onto proteins annotated with this EC number or GO term. As the four metagenomes had different total base-pair counts of sequencing data, the raw base-pair counts of EC numbers and GO terms were normalized to their base-pair counts per 10 Gbp of sequencing data. To identify EC numbers and GO terms with statistically significant differences between the phosphorus-rich and phosphorus-deficient soils, differential analysis was performed using the likelihood ratio test with Benjamini–Hochberg multi-comparison correction using the edgeR package[41]. EdgeR has been shown to provide good sensitivity under controlled false discovery rates for metagenomic comparisons[42] and stringent estimation of statistical confidence using only two replicates[43,44]. EC numbers and GO terms with a q value <0.05 and fold change ≥1.2 (refs 45,46) were considered to have statistically significant differences in their gene abundances. The gene abundance for an EC number from a family in the taxonomy tree was estimated as the normalized base-pair counts of reads mapped onto proteins annotated with this EC number from all genomes in this family. Differential analysis was performed as above to identify families with significantly different gene abundances between the phosphorus-deficient and phosphorus-rich soils for a given EC number (Fig. 2). Enrichment analysis was used to identify families enriched in genes with a given EC number. The enrichment analysis was conducted using Fisher's exact test with Benjamini–Hochberg multi-comparison correction across all families with a q value threshold of 0.05 (Fig. 2).

Assembly-based data analyses

The pre-processed reads from above were assembled on the Rhea supercomputer for each metagenome using Omega version 2.0 (ref. 37) with a minimum overlap length of 40 bp. Genes and proteins were called from the obtained scaffolds using the Prodigal algorithm[47] and annotated using the UniFam workflow[40]. The coverage depths of the scaffolds in the high- and low-molecular-weight DNA samples of a metagenome were estimated by mapping reads from the 2× 250-bp dataset and the 2× 150-bp dataset, respectively, using Bowtie 2 (ref. 48). Binning based on differential coverage

depths was performed using MetaBAT[49] and MaxBin[50]. Based on the binning quality evaluated by CheckM[51], MetaBAT was used to bin the metagenome of plot 6 and MaxBin was used for plots 1, 30 and 36. The obtained complete or near-complete genomes had >90% completeness and <10% contamination based on the CheckM results. The taxonomic classifications of these genomes were inferred from a phylogenetic tree constructed with reference genomes using PhyloPhlAn[52]. The phylogenetic tree was visualized using the Interactive Tree of Life online interface[53]. The metabolic network of novel genomes was reconstructed using UniFam[40]. The operons of the phytase genes were identified by requiring the same translation direction for all genes in an operon[54] and less than 50 bp of intergenic distance between neighbouring genes in an operon. Taxonomy classification for scaffolds containing phytase genes was inferred using Kraken[55]. The protein sequences of identified phytases were aligned with curated phytases from Swiss-Prot. The gene tree of phytase was generated using alignment results and visualized in the Interactive Tree of Life online interface[53].

Metaproteomic measurements

Proteins were extracted from the composite soil samples using the NoviPure Soil Protein Extraction Kit (Mo Bio) as described previously[17]. The following modifications were made to the manufacturer's protocol. The cell lysates were concentrated using Amicon Ultra-4 Centrifugal Filter Units (30 KDa molecular weight cut-off; Millipore), then proteins were precipitated by adding trichloroacetic acid and incubating overnight at 4 °C. Protein pellets were re-solubilized in guanidine (6 M) and dithiothreitol (10 mM). Bicinchoninic acid assays were conducted to estimate the protein concentrations of the extracted protein samples. 50 µg of protein from each soil sample was further processed with the filter-aided sample preparation protocol, as described previously[56]. Proteins were first trypsin-digested overnight in an enzyme-to-substrate ratio of 1:100 (weight:weight) at room temperature with gentle shaking, followed by a second digestion for 4 h. All digested peptide samples were stored at −80 °C.

Shotgun proteomics measurements were carried out with 11-step online multidimensional protein identification technology on an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific), as described previously[56]. In each multidimensional protein identification technology run, 25 µg of peptides were loaded offline into a two-dimensional back column with internal diameter of 150 µm (Polymicro Technologies) packed with 3 cm of C18 reverse-phase resin (Luna; Phenomenex) and 3 cm of strong cation exchange resin (Luna; Phenomenex). Each multidimensional protein identification technology run was configured with 11 strong cation exchange steps using 5, 7, 10, 12, 15, 17, 20, 25, 35, 50 and 100% of solvent D (500 mM ammonium acetate in 95% water, 5% acetonitrile and 0.1% formic acid). Each strong cation exchange fraction was separated by a 110 min reverse-phase gradient from 100% solvent A (95% water, 5% acetonitrile and 0.1% formic acid) to 50% solvent B (30% water, 70% acetonitrile and 0.1% formic

acid). Both mass spectra (MS) and tandem mass spectra (MS/MS) were acquired in Orbitrap with resolutions of 30,000 and 15,000, respectively. The top-ten most abundant precursor ions were selected for MS/MS analysis by higher-energy collisional dissociation (HCD) after each MS scan. Each metaproteome sample was measured in technical duplicate.

The acquired MS/MS data from a soil sample were searched using the Sipros algorithm[57,58] against the matched protein database constructed from the metagenome of this soil sample. The ion abundance of an identified protein was estimated using the ProRata algorithm[57]. The protein abundance of an EC number or GO term in a soil sample was estimated as the total ion intensity of all proteins annotated with this EC number or GO term and normalized across the four metaproteomes[56,59]. EC numbers and GO terms identified in only one sample out of the four were discarded for statistical comparison. Missing values for protein abundances of EC numbers and GO terms not identified in individual samples were imputed using the Perseus[60] algorithm from normal distributions of 0.3 of the observed data variances around the minimum values based on the 'missing not at random' assumption in label-free proteomics, as described previously[61]. The protein abundances of EC numbers and GO terms were compared between the phosphorus-rich and phosphorus-deficient soils using the likelihood ratio test with Benjamini–Hochberg multi-comparison correction in the edgeR package[41]. The EC numbers and GO terms with q values less than 0.05 were considered to have significantly different protein abundances.

Soil property measurements

Physicochemical properties were measured for soils collected from the four plots. The Buoycous hydrometer method was used to determine sand, silt and clay fractions[62]. The soils were not pre-treated to remove oxides or organic matter and, thus, the measured sand, silt and clay fractions represented an effective texture. Soil pH was observed after shaking a 2:1 solution:soil mixture of 5 mM $CaCl_2$ solution and soil for 1 h and measuring the pH of the supernatant after centrifugation. Total carbon and nitrogen were measured by combustion on a LECO TruSpec carbon/nitrogen analyser (LECO Corporation). Total iron, aluminium and manganese oxides were obtained by dithionite–citrate–bicarbonate extraction[63] and analysed using inductively coupled plasma MS (ELAN 6100; PerkinElmer). Microbial biomass carbon and microbial biomass nitrogen were measured by the chloroform fumigation–extraction method[64,65]. Dissolved carbon and nitrogen were determined through extraction with 0.5 M K_2SO_4 added at a 7:1 solution-to-dry-soil ratio, shaking for 1 h in a horizontal shaker at 200 rpm and filtration through Whatman grade 1 filter paper. All filtrates were analysed using a Shimadzu Total Organic Carbon Analyzer (model TOC-LCSH) and Total Nitrogen Analyzer (model TNM-L). Microbial biomass carbon was calculated assuming an extraction efficiency of 0.45 (ref. 64), and microbial biomass nitrogen was calculated assuming an extraction efficiency of 0.54 (ref. 65).

Fourier transform ion cyclotron resonance MS analysis of soil organic matter

Fourier transform ion cyclotron resonance analysis of soil organic matter was performed as described previously[66]. Briefly, soil samples were sequentially extracted using three solvents with decreasing polarity (water–methanol–chloroform), starting from 1 g of soil. The high-resolution mass spectra of the organic matter in the three extracts were collected in negative ion mode on a 12-Tesla Bruker SolariX (Bruker Daltonics) Fourier transform ion cyclotron resonance MS instrument outfitted with a standard electrospray ionization interface. Electrospray ionization MS data were acquired for the *m/z* range of 100 to 1,300 with a resolution of 4.5×10^5 at $m/z = 451$ and an ion accumulation time of 0.1 s. Spectra were calibrated by two internal series of dissolved organic matter homologous series separated by 14 Da (corresponding to $–CH_2$ groups) and the mass accuracy was calculated to be <1 ppm for singly charged ions. Molecular formula assignments were made using a modified version of the compound identification algorithm[67]. The fractions of different biochemical classes of compounds were calculated based on their hydrogen-to-carbon and oxygen-to-carbon atomic ratios, as described previously[68].

Enzyme assays

Enzyme assays were performed on fresh soils collected in February 2016 using the same soil collection procedure as described above for metagenomics and metaproteomics. Enzyme activities in the soil samples were measured on substrates labelled with 4-methylumbelliferone (MUB). The substrates of tested enzymes included 4-MUB-α-D-glucopyranoside for α-glucosidase, 4-MUB-β-D-glucopyranoside for β-glucosidase, 4-MUB-β-D-cellobioside for β-D-cellobiosidase, 4-MUB-*N*-acetyl-β-D-glucosaminide for *N*-acetyl β-glucosaminidase, 4-MUB-β-D-xylopyranoside for β-xylosidase, 4-MUB-phosphate for phosphomonoesterase and bis-4-MUB-phosphate for phosphodiesterase[3]. For each enzyme assay, 2.75 g (dry weight) of soil was mixed with 91 ml of 50 mM sodium acetate buffer with a pH adjusted to match the pH of the soil samples. The soil slurry was split into three technical replicates. Some 800 µl of the soil slurry was added into every well in 96-well plates. For standard samples, 200 µl of MUB solutions at concentrations of 2.5, 5, 10, 25, 50 and 100 µM were added to the soil slurry. For enzyme reaction samples, 200 µl of each MUB-tagged substrate (200 µM) was added to the soil slurry. All standard and enzyme reaction plates were then incubated at 26 °C for 3 h. The plates were centrifuged for 8 min at 1,500 rpm and 250 µl of the supernatant in each well was transferred to a reading plate. To terminate the reactions, 5 µl of 0.5 M NaOH was added to each well. The standard and enzyme reaction plates were measured for the absorbance at 365 nm on a Tecan Infinite M200 microplate reader. The standard curves were constructed from standard samples. The enzyme activities were quantified in the unit of nmol g dry soil^{-1} h^{-1}.

Data availability

The sequencing data that support the findings of this study are publicly accessible from the National Center for Biotechnology Information Bioproject database with the accession number PRJNA374886. The protein databases and identification results from metaproteomics were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005910.

References

Condron, L. M., Turner, B. L. & Cade-Menun, B. J. in *Phosphorus: Agriculture and the Environment* (eds Sims, J. & Sharpley, A.) 87–121 (American Society of Agronomy, Madison, 2005).

2. Sharma, S. B., Sayyed, R. Z., Trivedi, M. H. & Gobi, T. A. Phosphate solubilizing microbes: sustainable approach for managing phosphorus deficiency in agricultural soils. *Springerplus* 2, 587 (2013).

3. Turner, B. L. & Wright, S. J. The response of microbial biomass and hydrolytic enzymes to a decade of nitrogen, phosphorus, and potassium addition in a lowland tropical rain forest. *Biogeochemistry* 117, 115–130 (2014).

4. Quesada, C. A. et al. Soils of Amazonia with particular reference to the RAINFOR sites. *Biogeosciences* 8, 1415–1440 (2011).

5. Wright, S. J. et al. Potassium, phosphorus, or nitrogen limit root allocation, tree growth, or litter production in a lowland tropical forest. *Ecology* 92, 1616–1625 (2011).

6. Wurzburger, N. & Wright, S. J. Fine-root responses to fertilization reveal multiple nutrient limitation in a lowland tropical forest. *Ecology* 96, 2137–2146 (2015).

7. Santiago, L. S. et al. Tropical tree seedling growth responses to nitrogen, phosphorus and potassium addition. *J. Ecol.* 100, 309–316 (2012).

8. Mayor, J. R., Wright, S. J. & Turner, B. L. Species-specific responses of foliar nutrients to long-term nitrogen and phosphorus additions in a lowland tropical forest. *J. Ecol.* 102, 36–44 (2014).

9. Liu, L., Gundersen, P., Zhang, T. & Mo, J. M. Effects of phosphorus addition on soil microbial biomass and community composition in three forest types in tropical China. *Soil. Biol. Biochem.* 44, 31–38 (2012).

10. Cassman, N. A. et al. Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland. *Sci. Rep.* 6, 23680 (2016).

11. Jonasson, S., Michelsen, A., Schmidt, I. K. & Nielsen, E. V. Responses in microbes and plants to changed temperature, nutrient, and light regimes in the Arctic. *Ecology* 80, 1828–1843 (1999).

12. Rinnan, R., Michelsen, A., Baath, E. & Jonasson, S. Fifteen years of climate change manipulations alter soil microbial communities in a subarctic heath ecosystem. *Glob. Chang. Biol.* 13, 28–39 (2007).

13. Koyama, A., Wallenstein, M. D., Simpson, R. T. & Moore, J. C. Carbon-degrading enzyme activities stimulated by increased nutrient availability in Arctic tundra soils. *PLoS. ONE* 8, e77212 (2013).

14. Mack, M. C., Schuur, E. A., Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S. Ecosystem carbon storage in Arctic tundra reduced by long-term nutrient fertilization. *Nature* 431, 440–443 (2004).

15. Tveit, A., Schwacke, R., Svenning, M. M. & Urich, T. Organic carbon transformations in high-Arctic peat soils: key functions and microorganisms. *ISME J.* 7, 299–311 (2013).

16. Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212 (2015).

17. Butterfield, C. N. et al. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. *PeerJ* 4, e2687 (2016).

18. Xue, K. et al. Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. *Nat. Clim. Change* 6, 595–600 (2016).

19. Jorquera, M. A. et al. Identification of beta-propeller phytase-encoding genes in culturable *Paenibacillus* and *Bacillus* spp. from the rhizosphere of pasture plants on volcanic soils. *FEMS Microbiol. Ecol.* 75, 163–172 (2011).

20. Mullaney, E. J. & Ullah, A. H. The term phytase comprises several different classes of enzymes. *Biochem. Biophys. Res. Commun.* 312, 179–184 (2003).

21. Lim, B. L., Yeung, P., Cheng, C. & Hill, J. E. Distribution and diversity of phytate-mineralizing bacteria. *ISME J.* 1, 321–330 (2007).

22. Turner, B. L. et al. Seasonal changes and treatment effects on soil inorganic nutrients following a decade of fertilizer addition in a lowland tropical forest. *Soil. Sci. Soc. Am. J.* 77, 1357–1369 (2013).

23. Turner, B. L., Yavitt, J. B., Harms, K. E., Garcia, M. N. & Wright, S. J. Seasonal changes in soil organic matter after a decade of nutrient addition in a lowland tropical forest. *Biogeochemistry* 123, 221–235 (2015).

24. Turner, B. L., Wells, A. & Condron, L. M. Soil organic phosphorus transformations along a coastal dune chronosequence under New Zealand temperate rain forest. *Biogeochemistry* 121, 595–611 (2014).

25. Turner, B. L. & Engelbrecht, B. M. J. Soil organic phosphorus in lowland tropical rain forests. *Biogeochemistry* 103, 297–315 (2011).

26. Turner, B. L. Resource partitioning for soil phosphorus: a hypothesis. *J. Ecol.* 96, 698–702 (2008).

27. Funk, J. L. & Vitousek, P. M. Resource-use efficiency and plant invasion in low-resource systems. *Nature* 446, 1079–1081 (2007).

28. Cleveland, C. C. & Liptzin, D. C:N:P stoichiometry in soil: is there a “Redfield ratio” for the microbial biomass? *Biogeochemistry* 85, 235–252 (2007).

29. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton, 2002).

30. Zechmeister-Boltenstern, S. et al. The application of ecological stoichiometry to plant–microbial–soil organic matter transformations. *Ecol. Monogr.* 85, 133–155 (2015).

31. Kaspari, M. et al. Multiple nutrients limit litterfall and decomposition in a tropical forest. *Ecol. Lett.* 11, 35–43 (2008).

32. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, 1982).

33. Bloom, A. J., Chapin, F. S. & Mooney, H. A. Resource limitation in plants—an economic analogy. *Annu. Rev. Ecol. Syst.* 16, 363–392 (1985).

34. Allison, S. D. & Vitousek, P. M. Responses of extracellular enzymes to simple and complex nutrient inputs. *Soil. Biol. Biochem.* 37, 937–944 (2005).

35. Gleeson, S. K. & Tilman, D. Plant allocation and the multiple limitation hypothesis. *Am. Nat.* 139, 1322–1343 (1992).

36. Hurt, R. A. et al. Improved yield of high molecular weight DNA coincides with increased microbial diversity access from iron oxide cemented sub-surface clay environments. *PLoS ONE* 9, e102826 (2014).

37. Haider, B. et al. Omega: an overlap-graph de novo assembler for metagenomics. *Bioinformatics* 30, 2717–2722 (2014).

38. Varghese, N. J. et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* 43, 6761–6771 (2015).

39. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60 (2015).

40. Chai, J., Kora, G., Ahn, T. H., Hyatt, D. & Pan, C. Functional phylogenomics analysis of bacteria and archaea using consistent genome annotation with UniFam. *BMC Evolut. Biol.* 14, 207 (2014).

41. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).

42. Jonsson, V., Osterlund, T., Nerman, O. & Kristiansson, E. Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genom.* 17, 78 (2016).

43. Zhang, Z. H. et al. A comparative study of techniques for differential expression analysis on RNA-Seq data. *PLoS. ONE* 9, e103207 (2014).

44. Rapaport, F. et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-Seq data. *Genome Biol.* 14, R95 (2013).

45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57 (2009).

46. Mosier, A. C. et al. Elevated temperature alters proteomic responses of individual organisms within a biofilm community. *ISME J.* 9, 180–194 (2015).

47. Hyatt, D., LoCascio, P. F., Hauser, L. J. & Uberbacher, E. C. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28, 2223–2230 (2012).

48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).

49. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165 (2015).

50. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607 (2016).

51. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055 (2015).

52. Segata, N., Bornigen, D., Morgan, X. C. & Huttenhower, C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4, 2304 (2013).

53. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–W478 (2011).

54. Oyserman, B. O., Noguera, D. R., del Rio, T. G., Tringe, S. G. & McMahon, K. D. Metatranscriptomic insights on gene expression and regulatory controls in *Candidatus Accumulibacter* phosphatis. *ISME J.* 10, 810–822 (2016).

55. Davis, M. P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A. J. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63, 41–49 (2013).

56. Li, Z. et al. Diverse and divergent protein post-translational modifications in two growth stages of a natural microbial community. *Nat. Commun.* 5, 4405 (2014).

57. Wang, Y., Ahn, T. H., Li, Z. & Pan, C. Sipros/ProRata: a versatile informatics system for quantitative community proteomics. *Bioinformatics* 29, 2064–2065 (2013).

58. Guo, X. et al. Sipros ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btx601 (2017).

59. Li, Z. et al. Integrated proteomics and metabolomics suggests symbiotic metabolism and multimodal regulation in a fungal-endobacterial system. *Environ. Microbiol.* 19, 1041–1053 (2017).

60. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740 (2016).

61. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J. Proteome Res.* 15, 1116–1125 (2016).

62. Gee, G. W. & Or, D. in *Methods of Soil Analysis: Part 4—Physical Methods* SSSA Book Series 5.4 (eds Dane, J. H. & Topp, G. C.) 255–294 (Soil Science Society of America, Madison, 2002).

63. Loeppert, R. L. & Inskeep, W. P. in *Methods of Soil Analysis: Part 3 — Chemical Methods* SSSA Book Series 5.3 (ed. Sparks, D. L.) 639–654 (Soil Science Society of America, Madison, 1996).

64. Beck, T. et al. An inter-laboratory comparison of ten different ways of measuring soil microbial biomass C. *Soil. Biol. Biochem.* 29, 1023–1032 (1997).

65. Brookes, P. C., Landman, A., Pruden, G. & Jenkinson, D. S. Chloroform fumigation and the release of soil nitrogen: a rapid direct extraction method to measure microbial biomass nitrogen in soil. *Soil. Biol. Biochem.* 17, 837–842 (1985).

66. Tfaily, M. M. et al. Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal. Chim. Acta* 972, 54–61 (2017).

67. Kujawinski, E. B., Longnecker, K., Barott, K. L., Weber, R. J. M. & Kido Soule, M. C. Microbial community structure affects marine dissolved organic matter composition. *Front. Mar. Sci.* 3, 45 (2016).

68. Tfaily, M. M. et al. Advanced solvent based methods for molecular characterization of soil organic matter by high-resolution mass spectrometry. *Anal. Chem.* 87, 5206–5215 (2015).

Acknowledgements

This work was supported by Laboratory Directed Research and Development funding from Oak Ridge National Laboratory (ORNL). The authors

acknowledge R. Hurt of ORNL’s Biosciences Division for assistance with DNA extractions from tropical soils and J. Phillips of ORNL’s Environmental Sciences Division for soil characterization. The metagenomic sequencing was conducted by the US Department of Energy (DOE) Joint Genome Institute (JGI). The Fourier transform ion cyclotron resonance MS analyses were performed by the Environmental Molecular Sciences Laboratory (EMSL). The JGI and EMSL are DOE Office of Science User Facilities sponsored by the Office of Biological and Environmental Research. This research used resources of the Oak Ridge Leadership Computing Facility. The ORNL and JGI are supported by the Office of Science of the US DOE under contract numbers DE-AC05-00OR22725 and DE-AC02-05CH11231, respectively.