Land-use influences phosphatase gene microdiversity in soils

Andrew L. Neal,[1]* Maike Rossmann,[1†]
Charles Brearley,[4] Elsy Akkari,[1] Cervin Guyomar,[2‡]
Ian M. Clark,[2] Elisa Allen[3§] and Penny R. Hirsch[2]

[1]*Sustainable Agricultural Systems Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.*
[2]*Department of Agroecology, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.*
[3]*Computational and Systems Biology Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.*
[4]*School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR4 7TJ, UK.*

Summary

Phosphorus cycling exerts significant influence upon soil fertility and productivity – processes largely controlled by microbial activity. We adopted phenotypic and metagenomic approaches to investigate phosphatase genes within soils. Microbial communities in bare fallowed soil showed a marked capacity to utilise phytate for growth compared with arable or grassland soil communities. Bare fallowed soil contained lowest concentrations of orthophosphate. Analysis of metagenomes indicated *phoA*, *phoD* and *phoX*, and histidine acid and cysteine phytase genes were most abundant in grassland soil which contained the greatest amount of NaOH-EDTA extractable orthophosphate. Beta-propeller phytase genes were most abundant in bare fallowed soil. Phylogenetic analysis of metagenome sequences indicated the phenotypic shift observed in the capacity to mineralise phytate in bare fallow soil was accompanied by an increase in *phoD*, *phoX* and beta-propeller phytase genes coding for exoenzymes.

However, there was a remarkable degree of genetic similarity across the soils despite the differences in land-use. Predicted extracellular ecotypes were distributed across a greater range of soil structure than predicted intracellular ecotypes, suggesting that microbial communities subject to the dual stresses of low nutrient availability and reduced access to organic material in bare fallowed soils rely upon the action of exoenzymes.

Introduction

For soil microorganisms, acquiring sufficient phosphorus (P) to support essential cellular processes can be onerous. The high reactivity of the orthophosphate anion (PO_4^{3-}) dictates that solution PO_4^{3-} concentrations are typically less than 10 μM (Ozanne, 1980; Mengel and Kirkby, 1987; Raghothama, 1999; Frossard *et al.*, 2000). Not only is the solution concentration maintained at low levels, but sequestered P is effectively retained and difficult to access. P is partitioned between inorganic (P_i) and organic (P_o) soil pools. pH is critical for P_i speciation: at low pH, soil solution P_i concentrations are controlled by sorption of PO_4^{3-} ions to iron or aluminium oxides and oxyhydroxides (Parfitt, 1989) where protonated bidentate inner-sphere complexes dominate. At higher pH, P_i chemistry is dominated by precipitation reactions with calcium, forming the increasingly insoluble dicalcium phosphate, octocalcium phosphate and hydroxylapatite, and hydrated minerals with iron and aluminium (Shen *et al.*, 2011). For P_o which may represent between 30 and 80% of total soil P (Harrison, 1987), pH is less critical for chemical speciation, but indirect influences on the soil microbiota may result in varying capabilities to mineralise different organic forms. Labile P_o species include mono- and diesters as well as polyphosphate esters, however inositol hexa*kis*phosphates (phytates, $insP_6$) and phosphonates (containing a carbon-phosphorus bond) are less labile (Turner *et al.*, 2002; Condron *et al.*, 2005).

In plants, the main compound for storage of P_i is phytate, particularly abundant in cereal grains. Phytate is an extremely stable compound and release of phytate-P requires activity of specific enzymes – inositol phosphohydrolases, alternatively known as phytases. Phytases

hydrolyse phosphate from phytate in a stepwise manner, yielding products that may then be hydrolysed further by phosphatases. The rapid turnover of crop residues in soils either directly or from the addition of plant materials in, for example, manures results in significant incorporation of plant-associated P into microbial biomass: as much as 28% of P in *Medicago* residues in the soil microbial biomass within 7 days has been reported (McLaughlin *et al.*, 1988). Free phytate is rapidly mineralized in both marine and terrestrial environments (Suzumura and Kamatani 1995; Doolette *et al.*, 2010) and organisms capable of hydrolysing phosphate esters may therefore exert a significant influence upon P cycling in soil. However, phytate can accumulate in soil as a result of interactions with mineral and organic soil components which renders it resistant to enzymatic breakdown.

There are several distinct phytase families (Mullaney and Ullah, 2003). Of these, histidine acid phytase (HAPhy), protein tyrosine phosphatase-like cysteine phytase (CPhy) and beta-propeller phytase (βPPhy) have been shown to be active in prokaryotes (Lim *et al.*, 2007). Structural differences between the families dictate that each dephosphorylates a different site. Lower-order products of phytate dephosphorylation may also be hydrolysed by phosphomonoesterases such as PhoA, PhoD (which also exhibits phosphodiesterase activity) and PhoX (Cosgrove, 1980) which also acts upon other phosphate esters. Differences in the environmental distribution of the various phosphatase genes are apparent from sequence databases. The *phoD* gene distribution is rather cosmopolitan, being dominant in a wide variety of environments from marine, including hydrothermal vents, and terrestrial systems in association with plants and animals (Ragot *et al.*, 2015), however *phoX* is less widely distributed but more abundant in terrestrial settings than marine (Ragot *et al.*, 2017). For phytases, there is a clear distinction between HAPhy and CPhy genes which are present only in terrestrial environments and βPPhy genes which are found in both marine and terrestrial environments (Lim *et al.*, 2007)

Microorganisms rely on exoenzymes to degrade complex organic material into simple forms to facilitate nutrient acquisition. In common with many organic compounds in soil, within relatively short timescales P_o may become stabilized by incorporation into macro- and micro-aggregates (Lützow *et al.*, 2006) rendering physical contact between cells and P_o challenging. The combination of chemical complexity and spatial separation in soils means that understanding the distribution and abundance of exoenzymes is critical to development of more accurate models of nutrient cycling in soils.

Plants have a limited capacity to access phytate in soil (Brinch-Pedersen *et al.*, 2002). Microbial activity therefore exerts a significant influence upon the ability of plants to access P_o (Steffens *et al.*, 2009; Richardson and Simpson, 2011), particularly in low P soils, or low input agriculture. Reduced use of inorganic agricultural fertilizers in response to financial pressures or to reduce nutrient loss to groundwater, or increased use of manures may require efficient microbial re-mining of P to ensure plant health and productivity. It is therefore important to understand the diversity of phosphatase genes in soils and the influence of land-use strategies. We therefore established a study of phosphatase genes in soil from a long-term experiment to investigate the effect of land-use upon gene abundance, activity and diversity in soil. Since exoenzymes are likely to play an important role in the mineralisation of P_o in soil we interpreted our results in the light if bioinformatics-based predictions of enzyme subcellular localisation. We demonstrate that phytate is utilized as a source of P most readily by cultivable bacteria from soil maintained as bare fallow, and this is associated with greater proportions of PhoD, PhoX and βPPhy exoenzymes in the bare fallow soil than in grassland or arable soils. Notwithstanding this shift in abundance, there is a remarkable degree of genetic similarity across the soils despite the extreme differences in land-use.

Results and discussion

Phosphorus in Highfield soils

Soil P was assessed as NaOH-EDTA extractable P from the Highfield land-use treatments (Fig. 1). The greatest concentration of P was identified in grassland soils (661.7 ± 31.3 μg-P g^{-1}). After 50 years of minimal inputs, bare fallow soils still retained measurable P concentrations (235.0 ± 3.8 μg-P g^{-1}). Arable soils contained intermediate levels of extractable P (517.0 ± 12.6 μg-P g^{-1}). For all comparisons, we employed one-factor analysis of variance (ANOVA): where a significant effect ($\alpha = 0.05$) of land-use was identified, differences between land-use means were determined by Holm-Šídák *post*-test comparisons. ANOVA indicated a significant effect of land-use upon soil P concentration. Comparisons of land-use means indicated that P concentration in each soil was significantly different from the others. The general trend in P of grassland > arable > bare fallow matches those observed for other soil measures previously reported for Highfield soils including carbon and nitrogen content (Table 1).

Quantitative comparisons of NaOH-EDTA extractable orthophosphate and phytate in the soils using suppressed ion conductivity-HPLC showed a significant effect of management upon orthophosphate (Fig. 1A), with greatest concentrations in grassland and arable soils (279.4 ± 17.4, 252.6 ± 8.4 μg g^{-1} dry soil, respectively) which were not significantly different, in contrast to 105.7 ± 6.7 μg g^{-1} extracted from bare fallow soil (significantly less than extractable orthophosphate in either grassland or arable
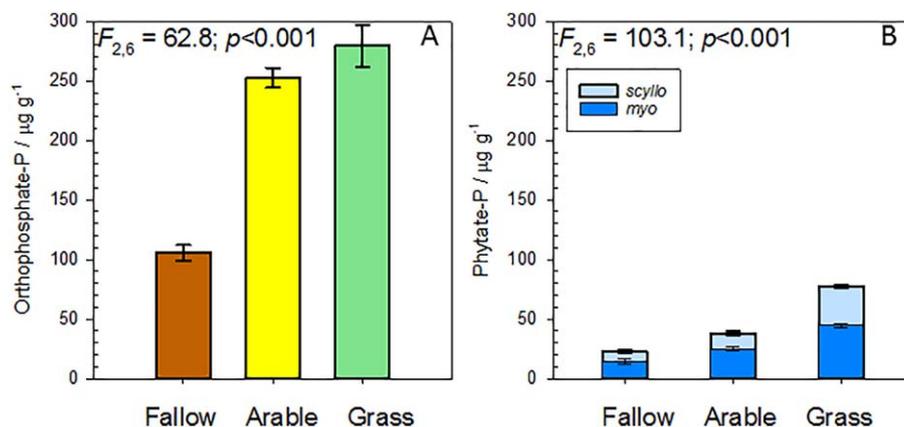
Fig. 1. Speciation of P in NaOH-EDTA extracts from Highfield soils.
A. There was a significant effect of land-use upon orthophosphate (PO_4^{3-}) concentrations in the three soils. There was no significant difference between the mean orthophosphate concentration in grassland and arable soils ($t = 2.01$; $p = 0.056$) but both soils had significantly higher concentrations than bare fallow soil ($p < 0.001$).
B. Phytate (IP_6) concentrations: only *myo*- and *scyllo*-stereoisomers were detected. There was a significant effect of land-use upon the total phytate concentration in the three soils. All mean phytate concentrations were significantly different from the other means ($p < 0.001$) Data was generated using suppressed ion conductivity HPLC of NaOH-EDTA soil extracts. In each case the mean of three estimates ± standard error of the means are shown.

soils). Greater differences were observed in the amounts of extractable phytate, where both *myo*- and *scyllo*-IP_6 stereoisomers were identified for all three contrasting soils (Fig. 1B). A significant effect of land-use was observed for total (*myo* + *scyllo*) phytate from the soils. The greatest amounts were extracted from grassland soils (77.1 ± 6.6 µg g^{-1}), compared with 38.1 ± 3.2 and 22.9 ± 2.8 µg g^{-1} extracted from arable and bare fallow soils, all mean estimates significantly different ($p < 0.001$). These values are typical of phytate estimates from a range of agricultural soils (Jarosch *et al.*, 2015).

Soil microbial community response to phytate

Growth rates of soil microbial communities were measured in a defined medium with phytate as the only added source of P. Based upon estimates of NaOH-extractable P, the maximum amount of autochthonous P added to the cultures was approximately 66 µg with grassland and 24 µg in bare fallow soils. These amounts are negligible compared with the amount of phytate added (2.94 mg) and are unlikely to have confounded the results. Preliminary

experiments indicated that where phytate was added to growth media, orthophosphate, assessed spectrophotometrically, accumulated to a greater extent in medium inoculated with bare fallow soil, apparently in response to cells entering stationary phase of growth (see Supporting Information Fig. S5). We observed highest microbial growth rates in cultures inoculated with bare fallow soil (Fig. 2A). Grassland soil communities appeared to be particularly ineffective in accessing the added phytate-P. In support of growth rate estimates, maximum likelihood (ML) MPN estimates based on data from the three replicate plots for each land-use type and using the same defined medium as used for growth estimates indicated significantly greater numbers of phytase-positive cultivable organisms in bare fallow soil (based upon the distribution of the ML 95% confidence intervals associated with the MPN estimates); arable and grassland soils were not significantly different (Fig. 2B). At the dilutions tested ($10^{-2} - 10^{-5}$-fold) there was no evidence of growth in dilution series of soils in the absence of added phytate: cells within replicates ($n = 8$ at each dilution) for each soil did not violate the assumption of non-random distribution (Grassland,

Table 1. Summary physical and chemical data of Highfield Ley-Arable experiment soils.

	Density[a]/ g cm^{-3}	pH[a] (H$_2$O)/-log (g[H$^+$]L^{-1})	Organic Carbon[a]/ mg g^{-1} soil	Free Organic Carbon[b]/ µg g^{-1} soil	Intra-aggregate Organic Carbon[b]/ µg g^{-1} soil	Nitrogen[a]/ µg g^{-1} soil	NaOH-EDTA extractable Phosphorus/µg g^{-1} soil
Grassland	0.99	6.0	3.9	4,690	3,010	390	662
Arable	1.30−1.45	5.8	1.3	370	490	150	517
Bare Fallow	1.30−1.45	5.1	0.8	150	380	100	235

a. Gregory *et al.* (2016).
b. Hirsch *et al.* (2009).
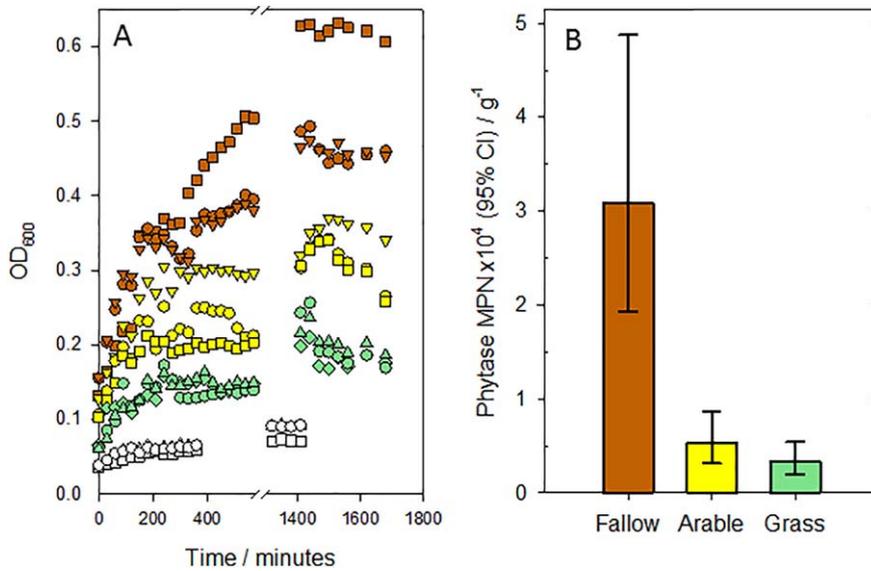
Fig. 2. Response of soil microbial communities to phytate.
A. Growth in replicate laboratory batch cultures inoculated with 100 mg g⁻¹ of air dried soil from each land-use in which phytate was the only added source of P. Grassland data points are shown in green, arable in yellow and bare fallow in brown. Within each land-use, individual replicate plots are represented by circles, squares and triangles. Growth in the absence of added phytate is shown by open data points representing the mean of grassland (circles), arable (triangle) and bare fallow (square) plots.
B. Maximum-Likelihood Most Probable Number estimates for phytase-positive cultivable organisms in the three land-use soils. Error bars represent 95% confidence intervals.

$\chi^2 = 2.61$, $p > 0.05$; Arable $\chi^2 = 7.31$, $p > 0.05$; Bare Fallow, $\chi^2 = 2.69$, $p > 0.05$). The observation that under conditions where phytate is the sole source of P, cultivable microbes in bare fallow soil are both more numerous (based upon ML-MPN estimates) and grow more quickly, are in marked contrast to other estimates of biological activity reported for these soils. Estimates of extractable ester-linked fatty acids and basal soil respiration (Gregory *et al.*, 2009), extractable DNA (Hirsch *et al.*, 2009) and extractable phospholipid fatty acids, biomass carbon and ATP content (Wu *et al.*, 2012) are all greatest in grassland soils and least in bare fallow soils with arable soils being intermediate. Thus, despite bare fallow soils having the lowest biomass, there were greater numbers of cultivable phytase-positive organisms which were able to grow more quickly and mineralize phytase more extensively in this study. This suggests a shift in the capacity of the community in bare fallow soil to exploit phytate as a P-source.

### Phylogenetic annotation of Highfield metagenome datasets

Soil from three replicate plots for each of the three land-uses were used to generate shotgun metagenome sequences. After filtering to remove substandard sequences, the average size of each metagenome was 356.7 million sequences (range 576.4–228.0 million) with an average sequence length of 100 nucleotide bases. Phylogenetic analysis of the complete datasets was performed using MEGAN and comparison against the RefSeq non-redundant protein database held at NCBI. Phylogenetic annotation of the sequences indicated that in each soil Actinobacteria, Alpha-proteobacteria and Bacteroidetes/Chlorobi accounted for approximately 50% of the sequences (Fig. 3A). SIMPER analysis of Bray-Curtis similarity

coefficients indicated a 12% dissimilarity across the averaged abundances for the three land-uses. Again using Bray-Curtis similarity coefficients, PERMANOVA analysis indicated a significant effect of land-use (*pseudo-F* = 4.34; $p = 0.004$ based upon 9,999 random permutations). A ternary plot of the relative distribution of orders/sub-phyla determined by MEGAN among the three land-uses (Fig. 3B) indicated that the Alpha-proteobacteria, Actinobacteria and Bacteriodetes were dominant in all three land-use soils but that members of the Gemmatimonadetes and Armatimonadetes were particularly associated with bare fallow soil, although at relative low abundance.
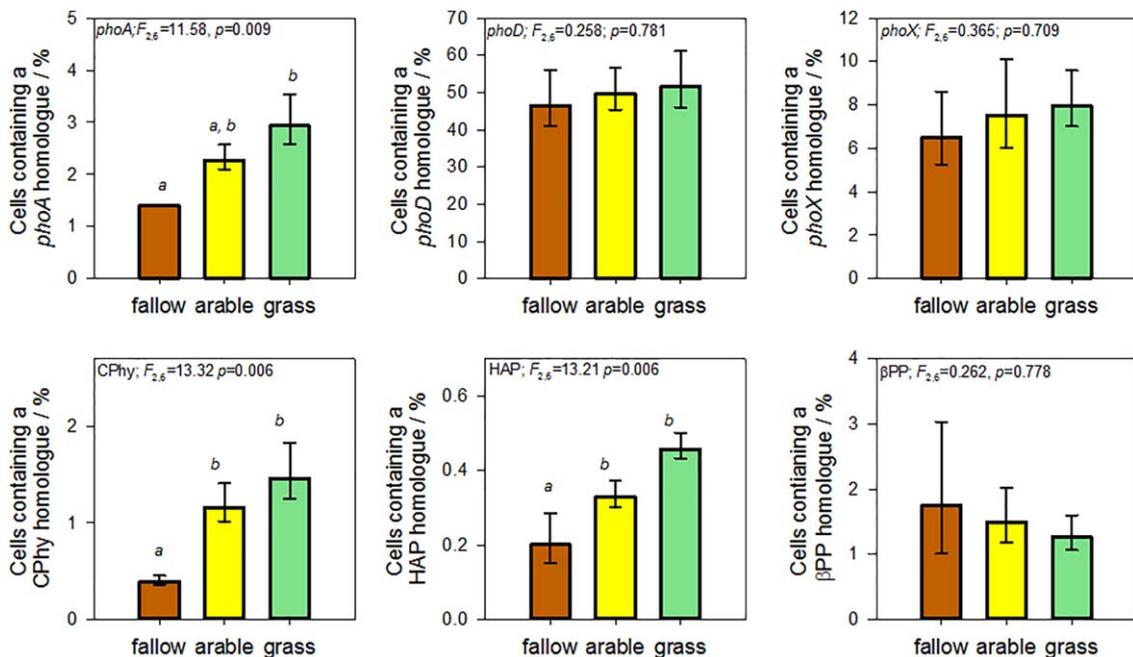
Abundance of phosphatase genes in Highfield soils

The bare fallow soil microbial community appears to be able to grow more efficiently on phytate compared with grassland or arable soils. We therefore interrogated metagenome sequence datasets generated from each soil for the presence of functional genes involved in dephosphorylation of organic P compounds. From the sequenced environmental DNA, we first calculated the mean abundance of four single-copy ubiquitous genes as an estimate of the number of genome-equivalents (GE) represented in each metagenome. There was no significant difference between the mean GE of 34,485 for grassland, 30,967 for arable and 33,646 for bare fallow soils ($F_{2,6}=0.06$; $p = 0.946$). The abundance of each phosphatase gene was compared against the estimates of GE for each metagenome dataset to determine the frequency (%GE) of each gene in each soil (details are given in SI3).

A comparison of homologue frequency across all genes (Fig. 4) indicates that phosphomonoesterases are more abundant in soil than phytases: *phoD* in particular is much more abundant than the other phosphatase genes. This is
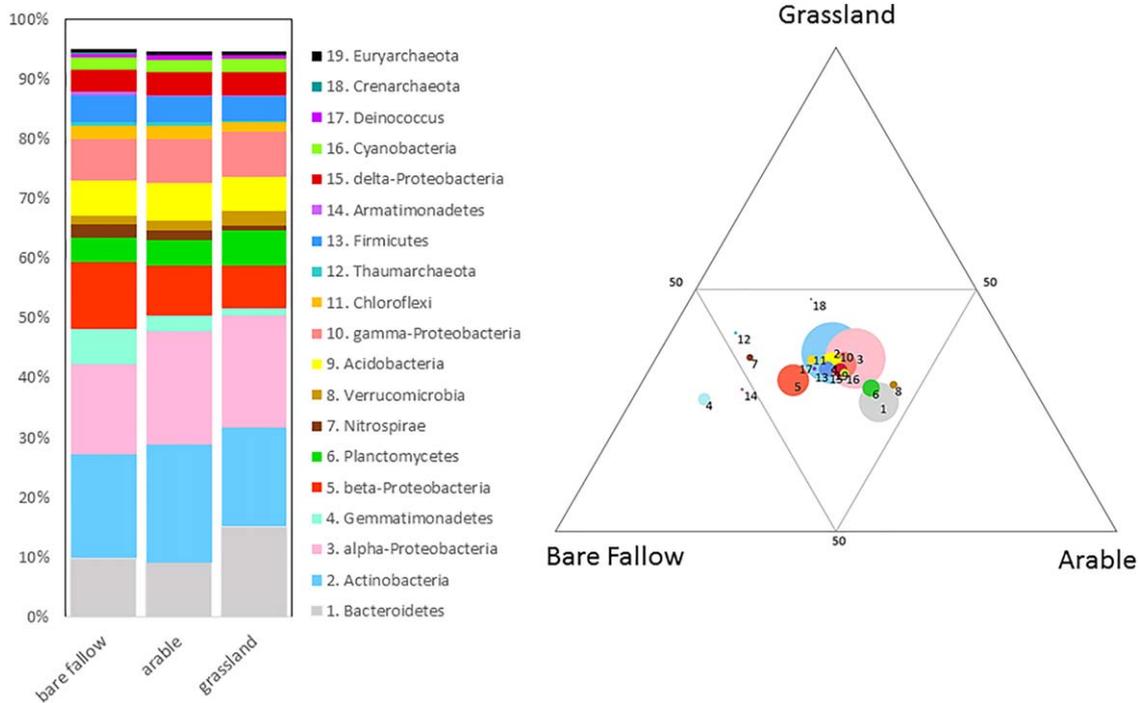
Fig. 3. (A) Relative sequence abundance of different sub-phyla in metagenomes generated from Highfield experimental soils. Actinobacteria, alpha-Proteobacteria and Bacteroidetes/Chlorobi accounted for approximately 50% of the sequences from all soils. (B) Ternary plot of the relative distribution of orders/sub-phyla determined among the three land-use soils. Alpha-proteobacteria, Actinobacteria and Bacteriodetes are dominant in all three soils but Gemmatimonadetes and Armatimonadetes are particularly associated with bare fallow soil, although they are present at relative low abundance.

Fig. 4. Normalized abundance of phosphatase gene homologous sequences identified in shotgun metagenome datasets generated from triplicate estimates for each land-use soil, error bars represent standard error. Counts of homologous sequences are normalized relative to the number of genome-equivalents in each metagenome, normalisation details are given in supplementary information. For each gene, the result of a one-factor analysis of variance is given. Where significant differences in normalized homologue abundance are identified, the results of Holm-Šídák *post*-test comparisons are given: different letters indicate significant differences in treatment means.

consistent with analysis of *phoA*, *phoD* and *phoX* abundance in publically available bacterial genomes and metagenome datasets (Tan *et al.*, 2013; Ragot *et al.*, 2015; Ragot *et al.*, 2017). In the case of the phosphomonoesterases (where grassland soil contained the greatest abundance), *phoD* represented 52%GE compared with 3%GE and 8%GE for *phoA* and *phoX* respectively. We could identify no significant effect of land-use upon *phoD* ecotype abundance (Fig. 4), contradicting previous studies which have shown that fertilisation causes changes in *phoD* abundance in agricultural soils assessed by quantitative PCR (Tan *et al.*, 2013; Fraser *et al.*, 2015). There was an effect of land-use upon *phoA* abundance; with significantly fewer homologous sequences identified from bare fallow than either arable or grassland soils.

PhoA is the least abundant alkaline phosphatase gene in soil – a similar distribution is observed in marine systems (Sebastian and Ammerman, 2009; Luo *et al.*, 2011). Nevertheless, *phoA* homologues were more abundant than phytases, where mean abundances were less than 2% of all prokaryotes. Of the phytases studied here, the HAPhy gene was the least abundant: CPhy and βPPhy genes were equally abundant in the soils, in contrast to marine systems where βPPhy is the dominant phytase gene (Lim *et al.*, 2007). The trend in gene homologue frequencies was for greater numbers in grassland than bare fallow soils, with abundance in arable soils being intermediate. βPPhy-homologous genes showed the reverse trend although differences between land-use were not significant. Thus, the most abundant phytase in soil was dependent upon land-use: CPhy was dominant in grassland soils, βPPhy in bare fallow soil: the genes were equally abundant in arable soil (Fig. 4). Collectively, the genes responded in one of two ways to land-use: *phoA*, CPhy and HAPhy gene homologue frequencies all responded significantly to land-use; *phoD*, *phoX* and βPPhy gene homologues showed no such response.

Phylogenetic distribution of phosphatase genes in Highfield soils

To identify the potential reasons for the two groups of genes' contrasting response to land-use, we used our MApPP workflow (SI3) to assess both abundance and phylogenetic structure of the community associated with each gene. All PhoA enzymes were predicted to be periplasmic by PSORTb; all metagenomes were dominated by sequences with homology to a clade of actinomycete *phoA* genes (cluster 1, Fig. 5). The proportion of metagenome reads associated with cluster 1 were greatest in grassland and least in bare fallow metagenomes (0.76%GE in grassland, 0.54% in arable soils and 0.28% in bare fallow soil). There was a significant effect of land-use upon the proportion of reads associated with this cluster ($F_{2,6} = 20.1$; $p = 0.002$) with all land-uses being significantly different from the others ($p = 0.024$). There was a second group of metagenome reads placed deeply within an Enterobacteriaceae clade (*phoA* cluster 2). The
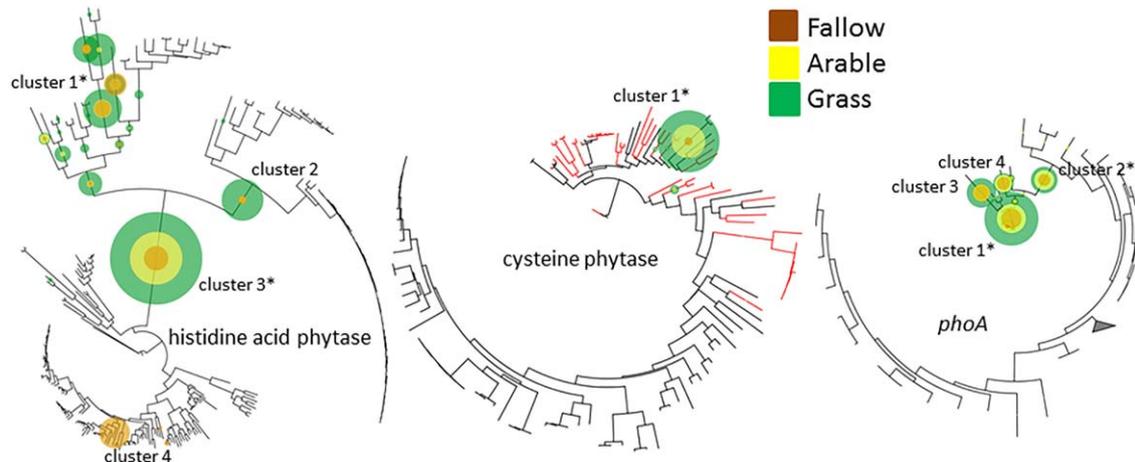
Fig. 5. Phylogenetic placement for metagenome sequences showing homology to the alkaline phosphatase *phoA* and histidine acid and PTP-like cysteine phytase genes. In each case there was a significant reduction in the normalized homologue abundance from grassland to bare fallow soils. Genes for which an exoenzyme is predicted are shown as red branches on the ML reference tree. Details of metagenome analysis are provided in the supplementary information. Histidine acid phytase cluster 1 represents gene sequences from *Yersinia, Pandoraea, Edwardsiella, Serratia* and *Burkholderia*; cluster 2, *Escherichia, Cronobacter* and *Shigella*; cluster 3 a group of sequences showing limited homology to clusters 1 and 2; and cluster 4, *Enterobacter* sp. strain 638. Cysteine phytase cluster 1 represents gene sequences from *Bdellovibrio* spp. PhoA cluster 1 represents a clade made up of *phoA* genes associated with *Streptoalloteichus, Amycolatopsis, Cellvibrio* and *Janibacter* species; cluster 2 represents a group of *phoA* genes from *Escherichia, Citrobacter, Providentia*, and *Shigella*; cluster 3, *Streptoalloteichus hindistanus*; and cluster 4, *Janibacter* sp. HTCC 2649. Clusters for which a significant effect of land-use was observed are indicated with *. Trees were generated using iTOL v3 (Letunic and Bork, 2016).

mean proportion of GE associated with cluster 2 were again greatest in grassland soil (0.36% GE), least in bare fallow soil (0.17%) and intermediate in arable soil (0.31%) with a significant effect of land-use ($F_{2,6}=20.5$; $p = 0.002$). *Post*-test comparisons indicated that the bare fallow soil exhibited a significantly reduced proportion than either the grassland ($t = 6.15$; $p = 0.003$) or arable ($t = 4.60$; $p = 0.007$) soils, but that there was no difference between the proportions for grassland or arable soils. Other placements show a similar trend in GE but no overall significant effect of land-use.

A number of CPhy proteins were predicted to be secreted: genes coding for these extracellular enzymes are shown as red branches in the reference phylogenetic tree (Fig. 5). Placement of metagenome reads for all soils was dominated by sequences assigned to the gene of the deltaproteobacterium *Bdellovibrio bacteriovorus* (CPhy cluster 1, Fig. 5), recently demonstrated to have a high degree of specificity for IP_6 (Gruninger *et al.*, 2014). The *B. bacteriovorus* CPhy enzymes were predicted to be extracellular. The number of metagenome reads with homology to the *B. bacteriovorus* CPhy were reduced in bare fallow soil (0.17% GE) compared with either arable (0.56%) or grassland (0.83%) soils and there was a significant effect of land-use ($F_{2,6}=15.97$; $p = 0.004$). *Post*-test comparisons indicated that bare fallow soil exhibited a significantly reduced proportion when compared with grassland ($t = 5.62$; $p = 0.004$) or arable ($t = 3.31$; $p = 0.032$) soils. There was no difference between the proportions for grassland and

arable soils. The only other placement of significant numbers of reads was a group showing homology to the CPhy gene associated with the Firmicute *Veillonella*.

The third gene for which significant reductions in abundance were observed was HAPhy. As with PhoA, HAPhy enzymes were predicted to be periplasmic. In this instance, placement of soil metagenome reads was dominated by two proteobacterial clades: (HAPhy clusters 1 and 2, Fig. 5). These phylogenetic associations were similar to that observed for *phoA*. There was a significant effect of land-use on the summed proportion of GE placed in cluster 1 (grassland 0.268%, arable 0.205%, bare fallow 0.129%; $F_{2,6}=8.17$; $p = 0.019$) but *post*-test comparisons indicated that only the difference between grassland and bare fallow was significant ($t = 4.04$; $p = 0.020$). Although a similar trend for greatest GE associated with grassland soil and the least for bare fallow soil was observed for cluster 2, there was no significant effect of land-use.

The largest placement of reads was at a deep branch within the tree (cluster 3, Fig. 5), suggesting dominant environmental HAPhy gene sequences may exhibit poor homology to sequenced genes. In all cases, the proportion of GE associated with placements was greatest in grassland soil, least in bare fallow soil (grassland 0.136%; arable 0.099%; bare fallow 0.026%): there was a significant effect of land-use ($F_{2,6}=56.82$; $p < 0.001$). Each land-use was significantly different from the others (smallest difference, grassland *versus* arable = 0.038%, $t = 3.61$; $p = 0.011$). Lastly, some metagenome reads showed high
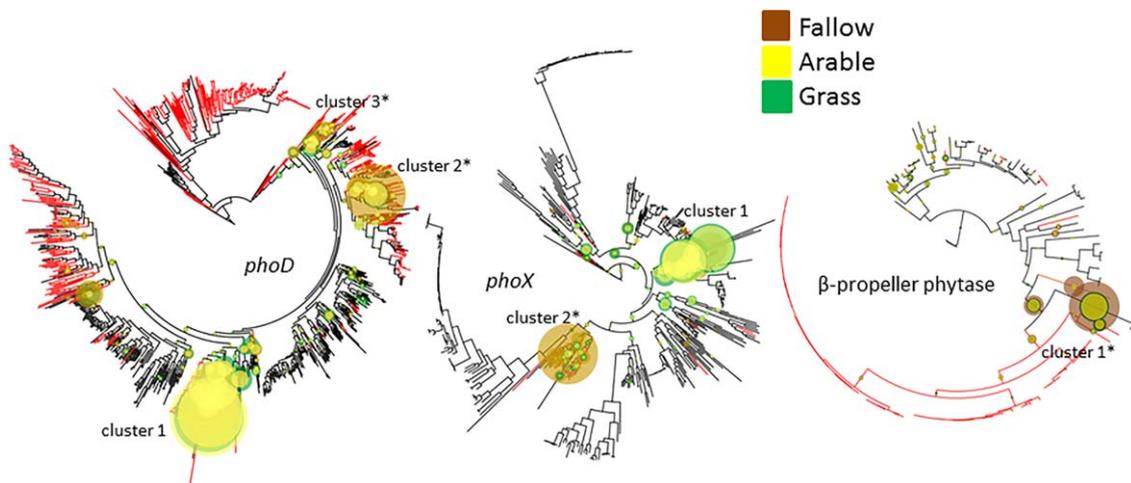
Fig. 6. Phylogenetic placement for metagenome sequences showing homology to the alkaline phosphatase *phoD* and *phoX* and β-propeller phytase genes. In each case there was no significant effect of land-use upon the normalized homologue abundance from each soil. Genes for which an exoenzyme is predicted are shown as red branches on the ML tree. Details of metagenome analysis are provided in the supplementary information. *PhoD* cluster 1 contains diverse sequences including *Cupriavidus*, *Ralstonia*, *Burkholderia*, *Rhodanobacter*, and *Candidatus* Koribacter; *phoD* cluster 2 represents sequences from *Variovorax*, *Acidovorax* and *Comamonas*; *phoD* cluster 3 represents sequences from *Gemmatimonas*, *Gemmatirosa*, Planctomyces and *Rhodopirellula*. *PhoX* cluster 1 represents genes from *Variovorax*, *Alcanivorax*, *Methylophaga*, and *Candidatus* Accumulibacter phosphatis; cluster 2 represents genes from *Acidovorax*, *Sphingomonas*, *Novosphingobium*, *Ramlibacter* and *Janthinobacterium*; β-propeller phytase cluster 1 represents phytase genes from *Bacillus*, *Paenibacillus*, *Alteromonas* and *Cyanothece*. Clusters for which a significant effect of land-use was observed are indicated with *. Trees were generated using iTOL v3 (Letunic and Bork, 2016).

homology to the HAPhy gene sequence of *Enterobacter* sp. strain 638 (HAPhy cluster 4, Fig. 5), an endophytic gammaproteobacterium with plant-growth promoting activity (Taghavi *et al.*, 2009). In this case, there were 20 metagenome reads showing homology from one sample from bare fallow soil, but none from any other sample.

In the case of the alkaline phosphatase PhoX, a number of enzymes were predicted to be extracellular (Fig. 6) although there is little correlation between phylogeny and extracellular localisation. In grassland soil, *phoX* homologues were most closely related to a proteobacterial clade (*phoX*, cluster 1, Fig. 6). The greatest number of reads for all soils within this clade showed high homology to the *phoX* sequence of the gammaproteobacterium *Methylophaga frappieri* with 0.499% GE from grassland metagenomes, 0.549% from arable metagenomes and 0.292% from bare fallow metagenomes. There was no significant effect of land-use upon the proportion of GE from each soil.

A distinct difference between the distributions of metagenome reads for *phoA*, CPhy and HAPhy homologues which showed similar phylogenetic distributions but a trend in reduced numbers of reads from grassland to bare fallow soil and *phoX*, is that a shift in phylogenetic distribution of reads from those associated with grassland and arable soils to those associated with bare fallow soil is evident. For bare fallow soil, the dominant placement (0.730% GE in metagenomes) showed high homology to the betaproteobacterium *Ramlibacter tataouinensis* TTB310 (*phoX* cluster 2, Fig. 6), originally isolated from semi-arid sandy soil (De Luca *et al.*, 2011). Only 0.111% GE and 0.168% grassland GE showed such homology. There was a significant effect of land-use ($F_{2,6}=5.95$; $p=0.038$), however *post*-test comparisons did not indicate any significant differences. A second set of reads within this cluster showed high homology to *phoX* of a second betaproteobacterium *Janthinobacterium lividum*. In this case, there were 0.120% GE from bare fallow metagenomes, 0.020% from arable metagenomes and 0.011% reads from grassland soil. There was a significant effect of land-use ($F_{2,6}=8.60$; $p=0.017$) and *post*-test comparisons indicated a significantly greater proportion in bare fallow than either grassland ($t=3.73$; $p=0.029$) or arable ($t=3.43$; $p=0.028$) metagenomes but no difference between grassland and arable soils. A subcellular localisation for either PhoX protein could not be assigned. However, both are part of a clade predicted to be extracellular, suggesting that bacteria dominating bare fallow soil express exoenzymes. The TTB310 genome possesses type II and III secretion systems – evidence that secreted hydrolases may be important in the desert sands where it originated (De Luca *et al.*, 2011).

For the dominant phosphatase gene, *phoD*, there is a clear correlation between phylogeny and subcellular localisation. A number of deeply rooting clades code for extracellular enzymes. The greatest number of placements was from arable soil, where 3.145% GE showed high homology to *phoD* of the gammaproteobacterium *Rhodanobacter spathiphylli* B39 (*phoD* cluster 1, Fig. 6). In comparison, this placement represented 2.38% GE of grassland metagenomes and 2.273% in bare fallow soil metagenomes but there was no significant effect of land-use. A closely related *phoD* homologue from the acidobacterium *Ca.* Koribacter versatilis was also most abundant in arable soil (2.41% GE) compared with grassland and bare fallow soils (1.60% and 2.00% respectively). Organisms harbouring B39 or *Ca.* K. versatilis *phoD* homologues (possibly also the *phoX* homologue of *Methylophaga frappieri*) appear well adapted to arable soil, however whether this was due to P-availability is unclear since phosphate fertilizer is added. B39 was isolated from the rhizosphere of the monocotyledon *Spathiphyllum* (De Clercq *et al.*, 2006) – it is possible that these *phoD*-expressing organisms are suited to the wheat rhizosphere. Other, more deeply placed sequences were also most abundant in arable soil. The B39 PhoD enzyme could not be assigned a subcellular compartment. The *Ca.* K. versatilis protein was predicted to be extracellular. These homologues are unique among the genes studied here, as the dominant placements are associated with arable soil. A second *phoD* clade dominates bare fallow soil (*phoD* cluster 2, Fig. 6). The largest numbers of metagenome reads in this clade had high homology to a Burkholderiales bacterium – JOSHI 001, 2.05% GE from bare fallow soil, compared with 1.06% and 1.09% from grassland and arable soils respectively. There was a significant effect of land-use ($F_{2,6}=24.17$; $p=0.001$) and the abundance was significantly greater in bare fallow soil than either grassland ($t=6.11$; $p=0.003$) or arable ($t=5.93$; $p=0.002$). Abundance in grassland and arable soils were not significantly different. PSORTb predicted PhoD of JOSHI 001 to be extracellular.

Within the same large clade, a number of metagenome reads showed high homology to *phoD* of *Gemmatirosa kalamazoonensis* KBS708, an agricultural soil isolate (De Bruyn *et al.*, 2013) and the related *phoD* sequence of polyphosphate-accumulating *Gemmatimonas aurantiaca* T-27 (*phoD* cluster 3, Fig. 6). The combined abundance of these two closely related sequences was greatest in bare fallow soils (0.961% GE) compared with arable (0.591%) and grassland (0.420%) and there was a significant effect of land-use upon the abundance ($F_{2,6}=7.71$; $p=0.022$) but *post*-test comparison indicated a significant difference only for grassland and bare fallow soils ($t=3.84$; $p=0.025$). PhoD proteins from both these strains were predicted to be extracellular. The observed shift in *phoD* community structure dependent upon land-use is consistent with previous studies based on PCR and DGGE (Tan *et al.*, 2013; Fraser *et al.*, 2015).

For βPPhy, there is a clear correlation between phylogeny and enzyme secretion, with a large phyletically-mixed clade (βPPhy cluster 1, Fig. 6) expressing exoenzymes and a second, more diverse clade with only a limited number expressing outer membrane-associated enzymes, the others being intracellular or of unknown localisation. The metagenomes for all soils were dominated by a set of reads showing homology to βPPhy from the cyanobacterium *Cyanothece* (βPPhy cluster 1, Fig. 6). For this placement, the greatest number of reads were identified in bare fallow soil (0.665% GE) with arable (0.283%) and grassland (0.264%) soils having a similar proportion of reads. There were also placements within this cluster showing high homology to the gammaproteobacterium *Alteromonas macleodii* (0.21% GE from bare fallow soil, 0.053% from grassland and 0.054% from arable). Of these two, a significant effect of land-use was only observed for *Cyanothece*-related abundance ($F_{2,6}=8.3$; $p = 0.019$) and *post*-test comparisons indicated that the bare fallow soil abundance was significantly greater than that for either the grassland ($t = 3.61$; $p = 0.033$) or arable ($t = 3.44$; $p = 0.027$) metagenomes. The *A. macleodii* βPPhy protein was predicted to be extracellular. Neither of the *Cyanothece* sp. βPPhy proteins could be assigned a subcellular compartment but were the only gene sequences placed within the clade which were not predicted to be extracellular. Other *Cyanothece* βPPhy proteins are predicted to be extracellular (*e.g.* B1WZU4_CYAA5 or A3ITE0_9CHRO), possibly indicating that the homologous sequences from these soils may code for extracellular phytase. Although this clade contains βPPhy sequences from *Bacillus* spp., few metagenomic sequences were detected suggesting that *Bacillus* phytases in sequence databases are relatively uncommon in soils.

Response of phosphatase gene expressing organisms in Highfield soils to land-use – Microbial communities in Highfield soils respond to land-use in different ways and display distinct gross phenotypes. Added phytate is mineralized more rapidly in soil fallowed for over fifty years compared with either arable soil receiving regular phosphate addition in fertilizers, or grassland soil. All measures of soil phosphate and P-containing compounds showed a consistent decrease from grassland to arable to bare fallow soils and previous studies indicate that organic matter and microbial abundance follow the same trend (Gregory *et al.*, 2009; Hirsch *et al.*, 2009).

In spite of this, MPN estimates for organisms capable of acquiring P from phytate were highest in bare fallow soil, evidence of a phenotypic shift in soil microbial communities. Analysis of phytase and phosphatase genes in metagenomes provides compelling phylogenetic evidence of this shift. For the genes *phoD*, *phoX* and βPPhy there were significant shifts in phylogeny, apparently towards organisms expressing exoenzymes in bare fallow soil.

Microbial secretion of exoenzymes has often been observed in response to low nutrient conditions or complex chemistries (Koch, 1985; Allison and Vitousek, 2005; Delpin and Goodman, 2009). The shift here is possibly in response to reduced P bioavailability in bare fallow soil. Organic matter in this soil is predominantly intra-aggregate (Hirsch *et al.*, 2009): not only do we measure reduced organic P in these soils, but access to it is likely to be more challenging because of occlusion within aggregates. Our observations are consistent with trait microdiversity – occurrence of ecologically distinct populations within phylogenetically-related groups (Zimmerman *et al.*, 2013) – suggesting microdiversity within phosphatase enzymes may be important in soil microbial communities and for soil fertility. However, not all observed responses of phosphatase gene homologues are necessarily driven by P availability. For instance, CPhy are dominated by sequences with homology to the CPhy gene of *Bdellovibrio bacteriovorus*, a predator of other bacteria. In this case the significant reduction of homologues from grassland to bare fallow soils is likely to be a direct response to reduced bacterial biomass observed for these soils (Hirsch *et al.*, 2009). Similarly, it is difficult to determine whether the increase in the cyanobacterium *Cyanothece* sp. βPPhy homologues in bare fallow soil is due to adaptation to low P availability – or indeed autotrophic metabolism more generally – or a response to reduced competition for light. It is also worth noting that the soils used in this study exhibited a progressive reduction in pH from grassland to bare fallow soil of almost a whole pH unit (Table 1). Soil pH has a clear effect upon both phosphomonoesterase and phosphodiesterase activities (Turner and Haygarth, 2005; Turner, 2010), some of which can be accounted for by varying pH optima of alkaline and acid phosphatases, and the community of *phoD*-expressing microbes (Ragot *et al.*, 2015).

To address these issues, responses of *phoD*, *phoX* and βPPhy ecotypes to soil chemical and structural factors were assessed using canonical correspondence analysis (CCA, Ter Braak, 1986). Counts of individual ecotypes represented in Fig. 6 acted as 'species' data: counts less than 5 were treated as 0 and ecotypes which were present in only 1 of the 9 soils were removed from the analyses. For each gene, ecotypes were divided into one of three categories based upon predictions of sub-cellular localisation by PSORTb: ecotypes which were predicted to be extracellular (shown in Fig. 6); ecotypes which were predicted to be intracellular; and ecotypes for which no localisation could be predicted. The pH, C_{org} and Nitrogen data presented in Table 1 as well as orthophosphate concentrations presented in Fig. 1 were used to represent soil chemical environments and the ratio of intra-aggregate C_{org} to free C_{org} (intra-aggregate ratio) – calculated from data presented in Table 1 – was used to represent soil
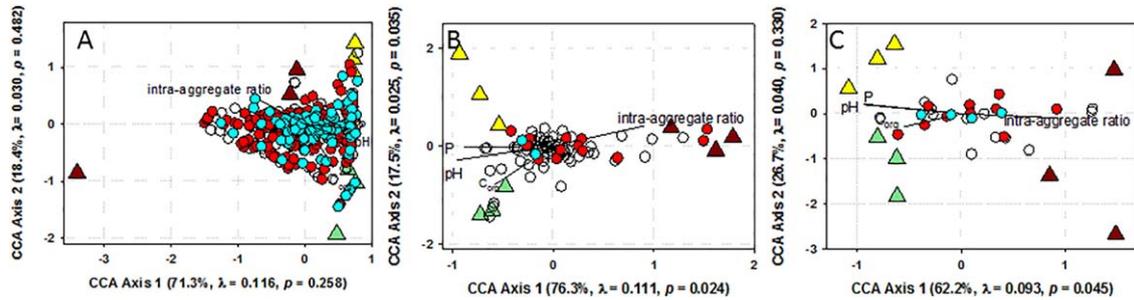
Fig. 7. Phylogeny-independent triplots of land-use, soil chemical and physical factors and *phoD* (A), *phoX* (B), and beta-propeller phytase (C) ecotypes associated with the Highfield Ley-Arable experiment. The plots are based upon CCA of gene ecotypes identified in soil metagenomes, shown in Fig. 6. Individual ecotypes are represented as circular data points and distinguished as coding for either extracellular (red) or intracellular (blue) proteins. Ecotypes for which no definite sub-cellular localisation could be predicted are shown as open circles. Triangles represent the centroids of individual replicate plots of grassland (green), arable (yellow) and bare fallow (brown) soils with regard to the ordination of relative frequencies across ecotypes. Environmental factors (pH, P, C_{org} and intra-aggregate ratio) are represented as vectors and increase in the direction of the vector: vector length indicates the degree of correlation of each environmental variable with ecotype relative frequencies. Values on each axis indicate the percentages of total variation explained by the axis, the eigenvalue (λ) and the permutation-based significance of the amount of total variation explained by the axis. Type 2 scaling is used in all cases.

physical structure. C_{org} and N were significantly correlated in the soils ($r = 0.999$, $p < 0.001$) and so only C_{org} was used in CCA. The variables were included in a separate model for each gene and model significance was estimated based upon 9,999 Monte Carlo permutations.

For *phoD*, although CCA clearly separated the land-uses (Fig. 7A), eigenvalues (λ) based upon the distribution of ecotypes between the three soil types associated with Axes 1 and 2 were not significant ($p > 0.05$). However, the median Axis 1 score for predicted extracellular ecotypes (–0.016, 95% confidence interval −0.094 to 0.042) was significantly lower than that for intracellular (0.142, 95% confidence interval 0.044 to 0.216) ecotypes (Mann-Whitney test, $U = 8122.5$, $p < 0.001$). For *phoX* (Fig. 7B), there was a separation of ecotypes on both Axes 1 ($\lambda = 0.111$, $p = 0.024$) and 2 ($\lambda = 0.025$, $p = 0.035$). To examine the environmental parameters responsible for ecotype groupings, scores were correlated with environmental parameters: both pH ($r = −0.959$, $p < 0.001$) and orthophosphate ($r = −0.882$, $p < 0.001$) were negatively correlated with Axis 1 ecotype scores and intra-aggregate ratio was positively correlated ($r = 0.915$, $p < 0.001$). There was no significant correlation of C_{org} with Axis 1 scores. Intra-aggregate ratio exerts a stronger effect than either pH and P for some of the PSORTb predicted extracellular *phoX* ecotypes; this is also the case for some of the ecotypes for which a subcellular location could not be predicted. There is an apparent distribution on Axis 2 which is associated with plant type, separating ecotypes between grass and arable (wheat) systems. The ordination of βPPhy gene ecotypes (Fig. 7C) is similar, although only Axis 1 was significant ($\lambda = 0.093$, $p = 0.045$). Again, ecotypes scores were negatively correlated with both pH ($r = −0.964$, $p < 0.001$) and orthophosphate ($r = −0.905$, $p < 0.001$), but positively correlated with intra-aggregate ratio ($r = 0.916$,

$p < 0.001$). Intra-aggregate ratio again appears to exert greater influence upon ecotype, especially those predicted to be extracellular, than either P or pH. Extracellular ecotypes of each gene predominate in soils low in P and pH but having high intra-aggregate ratios. It is difficult to separate potential chemical effects (pH or orthophosphate) from soil structural effects (intra-aggregate ratio) however, there is a common trend in the ordinations – ecotypes for which an extracellular mode is predicted are distributed across a greater range of intra-aggregate ratios than predicted intracellular ecotypes. This suggests that microbial communities subject to the dual stresses of low nutrient availability and reduced access to organic material in bare fallow soils rely more upon the production of exoenzymes (Allison, 2005; Allison *et al.*, 2011).

Irrespective of whether community shifts observed were driven exclusively by P-availability, soil structural effects, or a combination of both it is clear that land-use has the potential to induce shifts in the microbial community towards optimal states with benefits for nutrient cycling; this is particularly the case for the expression of exoenzymes by the microbial community. A salient observation from the CCA ordination of gene ecotypes is that the plant community itself (in this instance either mixed grasses and forbs, or wheat) appears to exert little structuring effect upon the distribution of functional genes. It is probably a combination of soil chemical and structural properties which exert the greater influence. A similar response of microbial communities has been observed across a broader range of soils (conventional and organic arable, pasture, natural grassland and deciduous forest) than used here (Kuramae *et al.*, 2011). In this case, pH and phosphate were identified, along with C:N ratio, as exerting significant influence upon total bacterial community structure.

Although the bare fallow soil studied here is extreme, it may be that periods of limited soil inputs maintain active populations of exoenzyme expressing organisms which are otherwise out competed in soils where nutrients are maintained in a high and freely available state. The dominant enzymes in the fallowed soil that we have demonstrated to overcome microbial P limitation will also make useful targets for heterologous expression in crop plant roots to improve crop P acquisition (*e.g.* George *et al.*, 2005; Chan *et al.*, 2006). It is also clear that in some cases, environmental genes associated with P-cycling are poorly represented in sequence collections and that in the case of the βPPhy well characterized enzymes such as *Bacillus* spp. do not appear to be well represented in the soils described in this study. Clearly, the wider relevance of changes in functional gene abundance and diversity remain uncertain in the absence of gene expression data. Since regulation of genes associated with the PHO regulon is tightly coupled to bioavailability of orthophosphate ions we chose not to study gene expression *via* metatranscriptomes generated from theses soils at this juncture since the timing of sampling will be crucial. However, the differences in gene abundance and especially the observed shift to exoenzymes in a land-used dependent manner warrant further study to link gene expression and phosphatase and phytase enzyme activity in these soils.

Experimental procedures

Soils

Experimental soils were collected from triplicate permanent grassland, arable and bare fallow plots (each of 50 × 7 m) of the Highfield Ley-Arable experiment (00:21:48W, 51:48:18N) at Rothamsted Research. The soil is a silty clay loam (27% clay) (Chromic Luvisol according to FAO criteria). At the time of sampling, arable plots had been under continuous arable (winter wheat, *Triticum aestivum* L., most recently Hereward seed coated with Redigo® Deter® combination insecticide/fungicide treatment, Bayer CropScience) and receiving ammonium nitrate fertilisation to provide approximately 220 kg-N ha^{-1} annum^{-1}, and additional 250 kg-K ha^{-1} and 65 kg-P ha^{-1} every three years for 62 years, bare fallow plots had been maintained crop- and weed-free by regular tilling for 52 years, and grassland plots had been maintained as a managed sward of mixed grasses for more than 200 years, mowed twice during summer months: all plots are considered now to be in quasi-equilibrium (Wu *et al.*, 2012). Physical and biological data has already been reported for these soils (Table 1).

Phosphorus chemistry in soils

We employed NaOH-EDTA extraction (Bowman and Moir, 1993) to estimate the amounts of orthophosphate and phytate in each soil. Total P in extracts was assessed by ICP-OES. Phytate and orthophosphate concentrations were assessed by suppressed ion conductivity HPLC (SIC-HPLC). Details are given in Supplementary Information 1 (SI1).

Growth and activity of soil microbial communities in response to added phytate

To assess the ability of microbial communities in the soils to mineralize and grow on phytate as the sole source of P (20 mM), we followed cell growth by optical density at 600 nm in triplicate, laboratory-based cultures in a phosphate-free defined mineral medium to which phytate was added. We also used a microwell-based approach (Rowe *et al.*, 1977) using the same medium to estimate the maximum likelihood most probable number (MPN) of cultivable phytase-expressing microorganisms calculated following a procedure described by Russek and Colwell (1983). We compared the number of wells showing growth at dilutions of $10^{-2} - 10^{-5}$-fold ($n = 8$ for each soil at each dilution) and included dilution series of soils and medium to which no phytate was added as a negative control. An independent experiment was performed for each land-use replicate ($n = 3$) and the data combined for MPN calculation. Details of culture media are given in SI2.

DNA extraction and metagenome sequencing

Soil was collected from triplicate plots for each land-use in October 2011 to a depth of 10 cm using a 3 cm diameter corer. The top 2 cm of soil containing root mats and other plant detritus was discarded. Ten cores per plot were pooled and thoroughly mixed whilst sieving through a 2 mm mesh; samples were then frozen at −80 °C. All implements were cleaned with 70% ethanol between sampling/sieving soil from each plot. Soil community DNA was extracted from a minimum of 2 g soil using the MoBio DNA PowerSoil® DNA isolation kit (Mo Bio Laboratories, Inc. Carlsbad, CA) with three replicates for each land-use soil. When necessary, extracts from separate 2 g aliquots from the same replicate were pooled to provide sufficient material for sequencing. 10 μg of high-quality DNA was provided for sequencing for each of the nine plots. Shotgun metagenomic sequencing of DNA from each soil was provided by Illumina® (Cambridge, UK) using a HiSeq™ 2000, generating 150bp paired-end reads. For analysis, sequences were limited to a quality threshold of 25 and minimum read length of 100bp using FASTX-Toolkit (version 0.0.13.2, http://hannonlab.cshl.edu/fastx_toolkit/index.html).

Metagenome analysis

To assess general abundance of organisms in the metagenomes, we mapped individual metagenomic sequences to the non-redundant (nr) protein database of NCBI (downloaded August 22nd 2016) using DIAMOND ver 0.8.27 (Buchfink *et al.*, 2015) in BLASTX mode and using a bitscore cut-off of 55. For each sequence, only the match with the highest bitscore was considered. Sequences not matching the nr database were considered currently unclassified. For classification of sequences with more than one nr entry with an equal bitscore, the weighted Lowest Common Ancestor (LCA) algorithm of MEGAN ver 6.5.8 (Huson *et al.*, 2007; Huson *et al.*, 2016) was employed using default parameters. The LCA

Table 2. Archetype proteins used to generate reference sets of proteins from the UniprotKB database.

Protein	UniprotKB Accession	Organism	Reference
Phosphatase			
PhoA	PPB_ECOLI	*Escherichia coli* K12	Bradshaw *et al.* (1981)
PhoD	ALPH_ZYMMO	*Zymomonas mobilis* subsp. *mobilis* ATCC 31821	Gomez and Ingram (1995)
PhoX	M7FUAO_VIBCL	*Vibrio cholera* O1	Majumdar *et al.* (2005)
Phytase			
AppA, Histidine acid phytase	PPA_ECOLI	*E. coli* K12	Greiner *et al.* (1993)
PhyL, β-propeller phytase	Q65NG0_BACLD	*Bacillus licheniformis* DSM 13	Kumar *et al.* (2014)
PhyA, PTP-like Cysteine Phytase	A3QMF6_9FIRM	*Mitsuokella multacida*	Kumar and Agrawal (2014)

algorithm assigns species-specific sequences to specific taxa. Sequences that are conserved across different species (*e.g.* as a consequence of horizontal gene transfer) are only assigned to taxa of higher rank (Huson and Mitra, 2012). Taxonomic comparison between the metagenomes was then made using MEGAN.

To look at the relative abundance and diversity of specific functional genes associated with organic P cycling we adopted an assembly-free, gene-centric approach, MApPP (Metagenomics/transcriptomics Assignment pHMM Phylogenetic Placement), to analysing the abundance and phylogenetic diversity of phosphatase genes associated with microbial communities in the different soils. Archetypal proteins for which positive activity has been demonstrated (Table 2) were used as a starting point to generate sets of reference proteins for each phosphatase. Details of MApPP analysis steps are provided in SI3. Subcellular localisation of each reference protein was predicted using PSORTb (Yu *et al.*, 2010: predictions available upon request to the corresponding author), the predictions are available upon request from the corresponding author. Nucleotide sequences corresponding to each protein sequence (available upon request from the corresponding author) were then used to query metagenome sequence datasets. To allow meaningful comparison between metagenomic datasets, gene abundance was expressed as a proportion of the estimated total number of genomes in each dataset, assessed by estimating the abundance of the ubiquitous, single-copy genes *rpoB*, *recA*, *gyrB* (Santos and Ochman, 2004) and *atpD* (Gaunt *et al.*, 2001). Nucleotide sequence-based pHMMs were developed for each gene as described above. Metagenome-derived homologue counts for each single-copy gene were size-normalized to the length of the shortest gene, *recA* accounting for differences in length between the genes. To do this, the modal length of *recA* (1,044 nt) was divided by the modal length of the other single-copy genes (1,422 nt for *atpD*, 2,415 nt for *gyrB*, 4,029 nt for *rpoB*), and this value was then multiplied by each single-copy gene count. The size-normalized abundance of each target phosphatase gene was then calculated for each soil as [target gene count·read length/(mean normalized counts of single-copy genes)] (Howard *et al.*, 2008). Metagenome sequences identified by our analyses have been deposited at the MG-RAST database (http://metagenomics.anl.gov/).

References

Allison, S.D. (2005) Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecol Lett* **8:** 626–635.

Allison, S.D., Weintraub, M.N., Gartner, T.B., and Waldrop, M.P. (2011) Evolutionary-economic principles as regulators of soil enzyme production and ecosystem function. In *Soil Enzymology*. Shukla, G., Varma, A. (eds). Berlin: Springer-Verlag, pp. 229–243.

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

Bowman, R.A., and Moir, O. (1993) Basic EDTA as an extractant for soil organic phosphorus. *Soil Sci Soc Am J* **57:** 1516–1518.

Bradshaw, R.A., Cancedda, F., Ericsson, L.H., Neumann, P.A., Piccoli, S.P., Schlesinger, M.J., *et al.* (1981) Amino acid sequence of *Escherichia coli* alkaline phosphatase. *Proc Natl Acad Sci USA* **78:** 3473–3477.

Brinch-Pedersen, H., Sørensen, L.D., and Holm, P.B. (2002) Engineering crop plants: getting a handle on phosphate. *Trend Plant Sci* **7:** 118–125.

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

Chan, W.L., Lung, S.C., and Lim, B.L. (2006) Properties of beta-propeller phytase expressed in transgenic tobacco. *Protein Expr Purif* **46:** 100–106.

Condron, L.M., Turner, B.L., and Cade-Menun, B.J. (2005). Chemistry and dynamics of soil organic phosphorus. In *Phosphorus: Agriculture and the Environment*. Sims, J.T., Sharpley, A.N. (eds). Madison WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 87–121.

Cosgrove, D.J. (1980). *Inositol Phosphates. Their Chemistry, Biochemistry and Physiology*. Amsterdam: Elsevier.

De Bruyn, J.M., Fawaz, M.N., Peacock, A.D., Dunlap, J.R., Nixon, L.T., Cooper, K.E., *et al.* (2013) *Gemmatirosa kalamazooensis* gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum *Gemmatimonadetes*. *J Gen Appl Microbiol* **59:** 305–312.

De Clercq, D., van Trappen, S., Cleenwerck, I., Ceustermans, A., Swings, J., Coosemans, J., *et al.* (2006) *Rhodanobacter spathiphylli* sp. nov., a gammaproteobacterium isolated from the roots of *Spathiphyllum* plants grown in a compost-amended potting mix. *Int J Syst Evol Microbiol* **56:** 1755–1759.

Delpin, M.W., and Goodman, A.E. (2009) Nitrogen regulates chitinase gene expression in a marine bacterium. *ISME J* **3:** 1064–1069.

De Luca, G., Barakat, M., Ortet, P., Fochesato, S., Jourlin-Castelli, C., Ansaldi, M., *et al.* (2011) The cyst-dividing bacterium *Ramlibacter tataouinensis* TTB310 genome reveals a well-stocked toolbox for adaptation to desert life. *PLoS One* **6:** e23784.

Doolette, A.L., Smernik, R.J., and Dougherty, W.J. (2010) Rapid decomposition of phytate applied to a calcareous soil demonstrated by a solution ^{31}P NMR study. *Eur J Soil Sci* **61:** 563–575.

Fraser, T., Lynch, D.H., Entz, M.H., and Dunfield, K.E. (2015) Linking alkaline phosphatase activity with bacterial *phoD* gene abundance in soil from a long-term management trial. *Geoderma* **257–258:** 115–122.

Frossard, E., Condron, L.M., Oberson, A., Sinaj, S., and Fardeau, J.C. (2000) Processes governing phosphorus availability in temperate soils. *J Environ Qual* **29:** 15–23.

Gaunt, M.W., Turner, S.L., Rigottier-Gois, L., Lloyd-Macgilp, S.A., and Young, J.P.W. (2001) Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J System Evol Microbiol* **51:** 2037–2048.

George, T.S., Simpson, R.J., Hadobas, P.A., and Richardson, A.E. (2005) Expression of a fungal phytase gene in *Nicotiana tabacum* improves phosphorus nutrition of plants grown in amended soils. *Plant Biotechnol J* **3:** 129–140.

Gomez, P.F., and Ingram, L.O. (1995) Cloning, sequencing and characterization of the alkaline phosphatase gene (*phoD*) from *Zymomonas mobilis*. *FEMS Microbiol Lett* **125:** 237–245.

Gregory, A.S., Watts, C.W., Griffiths, B.S., Hallett, P.D., Kuan, H.L., and Whitmore, A.P. (2009) The effect of long-term soil management on the physical and biological resilience of a range of arable and grassland soils in England. *Geoderma* **153:** 172–185.

Gregory, A.S., Dungait, A.J., Watts, C.W., Bol, R., Dixon, E.R., White, R.P., and Whitmore, A.P. (2016) Long-term management changes in topsoil and subsoil organic carbon and nitrogen dynamics in a temperate agricultural system. *Eur J Soil Sci* **67:** 421–430.

Greiner, R., Konietzny, U., and Jany, K.-D. (1993) Purification and characterization of two phytases from *Escherichia coli*. *Arch Biochem Biophys* **303:** 107–113.

Gruninger, R.J., Thibault, J., Capeness, M.J., Till, R., Mosimann, S.C., Sockett, R.E., *et al.* (2014) Structural and biochemical analysis of a unique phosphatase from *Bdellovibrio bacteriovorus* reveals its structural and functional relationship with the Protein Tyrosine Phosphatase class of Phytase. *PLoS One* **9:** e94403.

Harrison, A.F. (1987). *Soil Organic Phosphorus – A Review of World Literature*. Wallingford: CAB International.

Hirsch, P.R., Gilliam, L.M., Sohi, S.P., Williams, J.K., Clark, I.M., and Murray, P.J. (2009) Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. *Soil Biol Biochem* **41:** 2021–2024.

Howard, E.C., Sun, S., Blers, E.J., and Moran, M.A. (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* **10:** 2397–2410.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17:** 377–386.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., and Tappu, R. (2016) MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* **12:** e1004957. doi:10.1371/journal.pcbi.1004957

Huson, D.H., and Mitra, S. (2012). Introduction to the analysis of environmental sequences: metagenomics with MEGAN. In *Evolutionary Genomics: statistical and Computational Methods*. Anisimova, M. (ed). Berlin: Springer, pp. 415–429.

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

Koch, A.L. (1985). The macroeconomics of bacterial growth. In *Bacteria in Their Natural Environments*. Fletcher, M., and Floodgate, G.D. (eds). London: Academic Press, pp. 1–42.

Kumar, V., and Agrawal, S. (2014) An insight into protein sequences of PTP-like cysteine phytases. *Nusantara Bioscience* **6:** 102–106.

Kumar, V., Singh, G., Sangwan, P., Verma, A.K., and Agrawal, S. (2014) Cloning, sequencing, and *in silico* analysis of β-propeller phytase *Bacillus licheniformis* strain PB-13. *Biotechnol Res Int* **2014:** article ID 841353.

Kuramae, E.E., Yergeau, E., Wong, L.C., Pijl, A.S., van Veen, J.A., and Kowalchuk, G.A. (2011) Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol Ecol* **79:** 12–24.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44:** W242–W245.

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

Luo, H.W., Zhang, H.M., Long, R.A., and Benner, R. (2011) Depth distributions of alkaline phosphatase and phosphonate utilization genes in the North Pacific Subtropical Gyre. *Aquat Microb Ecol* **62:** 61–69.

Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Matzner, E., Guggenberger, G., Marschner, B., *et al.* (2006) Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions – a review. *Eur J Soil Sci* **57:** 426–445.

Majumdar, A., Ghatak, A., and Ghosh, R.K. (2005) Identification of the gene for the monomeric alkaline phosphatase of *Vibrio cholera* O1 strain. *Gene* **344:** 251–258.

McLaughlin, M.J., Alston, A.M., and Martin, J.K. (1988) Phosphorus cycling in wheat-pasture rotations. II. The role of the microbial biomass in phosphorus cycling. *Aust J Soil Sci* **26:** 333–342.

Mengel, K., and Kirkby, E.A. (1987). *Principles of Plant Nutrition*, 4th edn. Bern: International Potash Institute.

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

Ozanne, P.G. (1980). Phosphate nutrition of plants – general treatise. In *The Role of Phosphorus in Agriculture*. Khashawneh, F.E., Sample, E.C., and Kamprath, E.J. (eds). Madison WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 559–589.

Parfitt, R.L. (1989) Phosphate reactions with natural allophane, ferrihydrite and goethite. *J Soil Sci* **40:** 359–369.

Ragot, S.A., Kertesz, M.A., and Bünemann, E.K. (2015) *phoD* alkaline phosphatase gene diversity in soil. *Appl Environ Microbiol* **81:** 7281–7289.

Ragot, S.A., Kertesz, M.A., Mészáros, É., Frossard, E., and Bünemann, E.K. (2017) Soil *phoD* and *phoX* alkaline phosphatase gene diversity respond to multiple environmental factors. *FEMS Microbiol Ecol* **93:** fiw212.

Raghothama, K.G. (1999) Phosphate acquisition. *Annu Rev Plant Physiol Mol Biol* **50:** 665–693.

Richardson, A.E., and Simpson, R.J. (2011) Soil microorgansims mediating phosphorus availability. *Plant Physiol* **156:** 989–996.

Rowe, R., Todd, R., and Waide, J. (1977) Microtechnique for Most-Probable-Number analysis. *Appl Environ Microbiol* **33:** 675–680.

Russek, E., and Colwell, R.R. (1983) Computation of Most Probable Numbers. *Appl Environ Microbiol* **45:** 1646–1650.

Santos, S.R., and Ochman, H. (2004) Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ Microbiol* **6:** 754–759.

Sebastian, M., and Ammerman, J.W. (2009) The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J* **3:** 563–572.

Shen, J., Yuan, L., Zhang, J., Li, H., Bai, Z., Chen, X., *et al.* (2011) Phosphorus dynamics: from soil to plant. *Plant Physiol* **156:** 997–1005.

Steffens, D., Leppin, T., Luschin-Ebengreuth, N., Yang, Z.M., and Schubert, S. (2009) Organic soil phosphorus considerably contributes to plant nutrition but is neglected by routine soil-testing methods. *J Plant Nutr Soil Sci* **173:** 765–771.

Suzumura, M., and Kamatani, A. (1995) Mineralization of inositol hexaphosphate in aerobic and anaerobic marine sediments – implications for the phosphorus cycle. *Geochim Cosmochim Acta* **59:** 1021–1026.

Taghavi, S., Garafola, C., Monchy, S., Newman, L., Hoffman, A., Weyens, N., *et al.* (2009) Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Environ Microbiol* **75:** 748–757.

Tan, H., Barret, M., Mooij, M.J., Rice, O., Morrissey, J.P., Dobson, A., *et al.* (2013) Long-term phosphorus fertilisation increased the diversity of the total bacterial community and

the *phoD* phosphorus mineraliser group in pasture soils. *Biol Fertil Soils* **49:** 661–672.

Ter Braak, C.J.F. (1986) Canonical Correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67:** 1167–1179.

Turner, B.L., Papházy, M.J., Haygarth, P.M., and McKelvie, I.D. (2002) Inositol phosphates in the environment. *Philos Trans R Soc Lond B Biol Sci* **357:** 449–469.

Turner, B.L. (2010) Variation in pH optima of hydrolytic enzyme activities in tropical rain forest soil. *Appl Environ Microbiol* **76:** 6485–6493.

Turner, B.L., and Haygarth, P.M. (2005) Phosphatase activity in temperate pasture soil: potential regulation of labile organic phosphorus turnover by phosphodiesterase activity. *Sci Total Environ* **344:** 27–36.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., *et al.* (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26:** 1608–1615.

Wu, Y., Kemmitt, S., White, R.P., Xu, J., and Brookes, P.C. (2012) Carbon dynamics in a 60 year fallowed loamy-sand soil compared to that in a 60 year permanent arable or permanent grassland UK soil. *Plant Soil* **352:** 51–53.

Zimmerman, A.E., Martiny, A.C., and Allison, S.D. (2013) Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J* **7:** 1187–1199.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Fig. S1. Co-elution of *myo*- and *scyllo*-insP$_6$ standards with soil sample components. Fallow soil extract was analysed with or without added 2.5 μM *myo*- or *scyllo*-insP$_6$.

Fig. S2. Suppressed ion conductivity profiles of soil extracts. NaOH-EDTA extracts of soils managed as bare fallow, grassland and arable, diluted 10-fold with water and analysed by suppressed ion conductivity HPLC.

Fig. S3. Soil extract components are labile to *Escherichia coli* phytase. A fallow soil extract was analysed: without addition; after addition of 200 μM *myo*-insP$_6$; after addition of 200 μM *myo*-insP$_6$ and incubation with *E. coli* phytase; or after addition of enzyme alone. Orthophosphate resulting from dephosphorylation of insP$_6$ is identified by Pi. All samples were diluted 10-fold before HPLC.

Fig. S4. Separation of stereoisomers of inositol hexa*kis*phosphate. Aliquots (20 μL) of 10 μM inositol hexa*kis*phosphate were analysed by suppressed ion conductivity HPLC: D-*chiro*-, *myo*-, *neo*- and *scyllo*-insP$_6$.

Fig. S5. Response of soil microbial communities to phytate. Growth (bottom) and orthophosphate (P$_i$) accumulation (top) in laboratory batch cultures inoculated with 100 mg of air dried soil from each treatment in which phytate was the only added source of phosphorus. P$_i$ was assessed using spectrophotometric detection based upon the P$_i$ColorLock^{TM} Gold phosphate detection system (Innova Biosciences Ltd., Cambridge, UK) following the manufacturer's instructions.