Phosphohydrolase Gene Niche Separation in Soil and Maintenance of Microbiome Function under Organic and Inorganic Soil Fertilization.

Andrew L. Neal[1†] and Margaret J. Glendining[2]

[1]Department of Sustainable Agriculture Sciences and [2]Computational and Analytical Sciences, Rothamsted Research, Harpenden, Hertfordshire, UK.

†Corresponding author: andy.neal@rothamsted.ac.uk; Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ. UK. +44 (0)1582 763133

Abstract

The mechanisms by which microbial communities maintain functions within the context of changing environments are key to a wide variety of environmental processes. In soil, these mechanisms support fertility. Genes associated with hydrolysis of organic phosphoesters represent an interesting set of genes with which to study maintainance of function in microbiomes, since they participate in the same process and so in many respects are interchangeable. Here, we shown that the richness of ecotypes for each gene varies considerably in response to organic manuring and various inorganic fertilizer combinations . We show, at unprecedented phylogenetic resolution, that phylogenetic diversity of phosphohydrolase genes are more responsive to soil management and edaphic factors than the taxonomic biomarker 16S rRNA gene. Available phosphorus exerted no significant influence on gene distribution: instead we observed gene niche separation according to soil pH and exchangeable calcium. We infer a degee of competition between genes, ensuring that a gene most optimally adapted to the prevailing edaphic factors spreads through the population, thus maintining microbiome function.

Introduction

Global-scale censuses of soil microbiomes are identifying distinct patterns in the distribution of both bacterial species and community functions. A relatively small subset of bacterial phylotypes dominate soils across the globe[1] and the taxonomic and gene functional diversity of bacterial assemblages peak at mid-latitudes, declining towards the poles and equator[2]. Environmental variables such as climate (aridity, maximum temperature, precipitation characteristics), plant productivity, but especially soil pH[1-4] are more important than dispersal in determining global microbiome species assemblage and functions. The limited number of dominant phylotypes cluster into predictable ecological groups which share similar environmental niches[1], but habitat preferences are not linked to phylogeny at coarse biological resolution. It is also evident that the functional potential of soil microbiomes is enormously vast, but under-explored: most soil bacterial phylotypes are rare, relatively few abundant. Given this characteristic complex mix of numerous rare and few abundant phylotypes, the mechanisms by which important functions are maintained within microbiomes across ranges of climatic and edaphic factors, especially those brought about by land management or climate change, are both intriguing and not particularly well understood.

Understanding the effects of land management upon soil microbial communities is critical. Soil microbes are responsible for the provision of a significant number of environmental services including the regulation of biogeochemical cycles and delivery of nutrients to primary producers, degradation of pollutants and provision of clean water, regulation of atmospheric trace gases, and pest and pathogen control[5,6]. Microbial processes also play a large part in determining whether carbon (C) is released to the atmosphere as CO_2 or CH_4, or retained in soil[7], where it contributes to soil structural properties[8]. However, the molecular interactions between microbial species and their environment which influence the fate of soil C and other nutrients are largely unknown. In contrast to other habitats, microbial communities and substrates in soil are diverse and liable to physical protection and separation. Nutrients enter soil in both organic and inorganic forms. Organic forms (soil organic matter or SOM) are processed by the microbiome using energy derived largely from C in SOM. This C fuels a cascade of metabolic reactions that ultimately return chemically complex SOM to simple forms available for uptake by the microbiota and plants. The accessibility of this energy source depends on its nutritional context in organic matter (since complex bonds may have to be broken

to release it) and by the degree to which soil affords the molecules physical protection (in small pores or through chemical bonding to surfaces). Therefore, soil type and the nutritional complexity of inputs affect nutrient cycling rates and the fate of C and other nutrients in soil.

Organic nitrogen (N), sulfur and phosphorus (P) compound constituents of SOM are also degraded and assimilated as nitrate or ammonium, sulfate and phosphate respectively. Analagous to C processing, microbial processes play a central role in determining whether N in SOM is released to the atmosphere as N_2O or N_2 or retained in the soil. Fertility of soil depends to a large degree upon cycling of complex organic compounds to simple inorganic ions by the soil microbiome. However, modern agriculture has become dependant upon inorganic fertilizer to support ever greater crop yields - often at the expense of nutrient use efficicency and wider environmental pollution. Unlike C or N, P cannot typically be lost from soil to the atmosphere, but may be lost *via* run-off to groundwater or surface water bodies where it is the principal cause of eutrophication. The importance of organic phosphorus (P_{org}) in the global P-cycle, and the role that bacteria play in its cycling, has interested researchers since the beginning of the 20th century[9-11]. This is because of its importance in regulating movement of P between ecosystems and as a potential nutrient source to support plant, and particularly, agricultural production.

Thorough descriptions of the biodiversity of genes involved in P_{org} cycling have only recently been published (for the alkaline phosphatases *phoA, phoD* and *phoX*, refs 12-14 - bacterial non-specific acid phosphatases (NSAP), refs 15, 16 - phytases refs 14, 17) and the rest remain poorly described. Collectively, the group of gene families associated with hydrolysis of P_{org}-containing compounds represent an interesting set of genes with which to study how microbiomes maintain important functions in the face of environmental change. They all participate in the same process (acquisition of orthophosphate from P_{org}) so in many respects are interchangeable, and as genes coding for the catalysis of extracellular nutrients - and so positioned on the periphery of metabolic networks - are likely to be subject to horizontal genetic transfer between individual cells within microbiomes[18,19]. There are clear and consistent links between soil structural parameters and the abundance and diversity of bacterial genes coding for orthophpsphate-releasing exoenzymes. In directly comparable soils of contrasting SOM and structure, genes coding for intracellular phosphatase protein families

decrease in abundance in response to reduced soil structure while genes coding for protein families of both endo- and exoenzymes (alkaline phosphatases PhoD, PhoX, β-propeller phytase and class C non-specific acid phosphatase) show no reduction in abundance. Furthermore, for the latter protein families, genes predicted to code for exoenzymes are more abundant in poorly structured soil[14,16]. How diversity of these genes relates to species diversity, typically assessed using the 16S rRNA gene, remains unknown. To address this issue, we compared the abundance and phylogenetic diversity, using metagenomics, of a suite of phosphpohydrolase and 16S rRNA genes in soil from a long-term field experiment, comparing soil where fertility is managed organically to soils managed by the addition of various combinations of inorganic fertilizer. We show that the response of the phosphphydrolase genes to soil management is different from the response of the 16S rRNA gene and that individual phosphohydrolase genes occupy distinct niches within the soils, demarcated by soil pH and exchangeable calcium.

Results

Soil chemistry and phosphorus balance in Broadbalk soils – We compared four treatments on the Broadbalk winter wheat long-term experiment whose fertility is managed in contrasting ways. The four plots were: plot 2.2 to which farmyard manure (from cattle) has been added at a rate of 35 t ha^{-1} per year since 1843 (hereafter referred to as manured); plot 8 to which inorganic fertilizer containing 144 kg ha^{-1} nitrogen (N) since 1986 as Nitram® ammonium nitrate, 35 kg ha^{-1} phosphorus (P) as triple superphosphate (calcium dihydrogen phosphate), 90 kg ha^{-1} potassium (K) as potassium sulfate, and 12 kg ha^{-1} magnesium (Mg) as kieserite per year (referred to as fertilized) has been added since 1852; plot 5 lacking N-fertilization but receiving 35 kg ha^{-1} P, 90 kg ha^{-1} K and 12 kg ha^{-1} Mg per year (referred to as fertilized^{-N}) since 1852 and plot 20 lacking P-fertilization but receiving 192 kg ha^{-1} N, (96 kg ha^{-1} N 1906-2000) 90 kg ha^{-1} K and 12 kg ha^{-1} Mg per year since 1906 (referred to as fertilized^{-P}). Mean estimates of soil parameters are shown in Table I. Significant treatment effects were observed for most parameters: only soil pH (which is adjusted by application of calcium carbonate to maintain soil pH at a level which does not limit wheat yield), C/N ratio and exchangeable sodium (Na_{ex}) concentration showed no statistically significant differences associated with fertilization treatment. The highest concentrations of SOC (2.9 %), N (0.28 %), exchangeable potassium (K_{ex}) (610 mg kg^{-1}) and exchangeable magnesium (Mg_{ex}) (117 mg kg^{-1}) were recorded in manured soil. For other parameters higher values were observed for other treatments, fertilizer^{-P} in the case of exchangeable calcium (Ca_{ex}) (6.6 g kg^{-1}) and fertilizer in the case of gran yield, used here as a proxy for competition for P with the wheat crop (5.5 t ha^{-1}), however there was no statistically significant difference between those soils and manured soil (6.1 g kg^{-1} and 5.3 t ha^{-1} respectively).

Olsen-P concentrations have been recorded in the soils since 1865. Estimated Olsen-P in the original soils was low, at approximately 10 mg kg^{-1}, based upon measurements made on near-by plots in 1856. Up until 2000, Olsen-P increased progressively in manured, fertilized and fertilized^{-N} plots (Figure 1A) to over 80 mg kg^{-1}. At this point a decision was taken to cease additions of triple superphosphate fertilizer to the fertilized and fertilized^{-N} treatments with the result that Olsen-P in these soils has reduced consistently year on year. Measurement of Olsen-P on the fertilized^{-P} soil was only instigated in 1966 but has remained consistently below the estimated starting Olsen-P of 10 mg kg^{-1}. The highest Olsen-P concentrations of 106 mg kg^{-1} are observed in manured soil: the least, 3 mg kg^{-1}, in fertilized^{-P} soil. Over this same period, the P balance in the soils has varied greatly (Figure 1B). Up until 2000, manured, fertilized and fertilized^{-N} soils gained varying amounts of P each year. In contrast, fertilized^{-P} soil showed a consistent P-deficit, although this has gradually approached a zero balance as the wheat crop has yielded progressively less over time. Following the withdrawal of super triple phosphate from fertilized and fertilized^{-N} soils in 2000, both treatments now have negative P balances. Consequently,only manured soil currently has a positive P balance of 28 kg ha^{-1}, statistically significantly different from all other treatments which have negative balances ranging between -1 kg ha^{-1} for fertilized^{-P} soil and -14 kg ha^{-1} for fertilized soil.

Community response to treatments based upon 16S rRNA gene assemblage – Rarefaction of phylogenetic diversity (the sum of lengths of branches in a phylogenetic tree associated with metagenomic reads) based upon the 16S rRNA gene (Figure 2A) indicated that in no case was the complete 16S rRNA gene diversity captured, but the extent of diversity accounted for by sequencing was consistent between treatments. No significant differences were detected in balance-weighted phylogenetic diversity ($BWPD_1$) of the molecular marker 16S rRNA gene between treatments (Figure 2B). However, edge-PCA (Figure 2C) showed clear separation of treatments. The primary axis separated manured soil from soils receiving inorganic fertilization, particularly those lacking N- and P-fertilization. On this axis, organisms such as the δ-proteobacteria *Haliangium ochraceum* and *Steroidobacter denitrificans*, the verrucomicrobium *Candidatus* Xiphinematobacter sp. and the planctomycetes *Gemmata* sp. and *Phycisphaera mikurensis* were more abundant in manured soil while actinobacterium *Conexibacter woesei*, the chloroflexi *Caldilinea aerophila* and *Sphaerobacter thermophilus*, and *Gemmatimonas aurantiaca* and the closely related *G. phototrophica* were more abundant in fertilized^{-N} and fertilized^{-P} soils. On the second axis, fertilized soil was separated from all other treatments. *Ca*. Xiphinematobacter sp. was more abundant in fertilized soil, while *C. aerophila*, *S. denitrificans*, *Gemmata* sp. and *P. mikurensis* all had reduced abundance. These placements and difference sin the abundance of each placement can be seen in the phylogenetic placement shown in Figure 2D.

Phosphohydrolase gene phylogenetic diversity and response to soil fertilization – Comparison of rarefaction curves generated for each gene (Figure 3) indicated that in no case was the genetic diversity in the sampled communities represented completely. However, for each gene the extent of diversity accounted for by sequencing was comparable. These curves also demonstrated distinct differences in the abundance and phylogenetic diversity of the different genes. Alkaline phosphatase genes *phoD* and *phoX* were the most abundant and diverse phosphohydrolase genes in the soils. The two classes of NSAP studied presented similar phylogenetic diversity, although class C genes were more abundant. The remaining genes – the alkaline phosphatase *phoA* and the three phytase classes - all showed low diversity: only the β-propeller phytase (βPPhy) gene matched the abundance of the NSAP genes.

In contrast to shifts in microbial communities in response to soil fertilization evident from 16S rRNA phylogeny, where the dominant difference was in response to organic inputs from cattle manure *versus* inorganic fertilization, for most genes associated with phosphorus acquisition the major difference was between fertilized soil and the other treatments. This was most evident for the alkaline phosphatase *phoD* and class A NSAP. Also, except for acidic phytase genes, significant differences in $BWPD_1$ were identified for phosphohydrolase genes.

Alkaline phosphatase genes – No significant treatment effect on *phoD* normalized abundance was detected (Figure 4A), however gene $BWPD_1$ (Figure 4B) was significantly greater in soil receiving complete fertilizer than under the other treatments (smallest difference, $Q = 5.7$; $p = 0.016$). Genes in manured soil were also significantly more diverse than in fertilized^{-P} soil ($Q = 4.9$; $p = 0.033$). These differences were evident in edge-PCA ordination (Figure 4C) of soils using *phoD* phylogenetic placement (Figure 4D). Differences between fertilized and the other soils were distributed on the primary edge-PCA axis which accounted for over 90% of the total variability. The dominant ecotype in all soils showed homology to the *phoD* gene of *Parvularcula bermudensis* HTCC2503, however ecotypes with homology to genes of the γ-proteobacterium *Rhodoferax ferrireducens* T118, and β-proteobacteria *Rhodanobacter spathiphylli* B39 and *Limnobacter* sp. MED105 were more abundant in fertilized soils, while ecotypes with homology to genes of the actinobacteria *Streptomyces bingchenggensis* BCW1, *Ilumatobacter coccineus* YM16304 and *Ca*. Microthrix parvicella, and the γ-proteobacterium *Stenotrophomonas maltophila* strains were more abundant in the other soils. On the second edge-PCA axis, manured soil was separated from fertilized^{-P}. On this axis, ecotypes with homology to *S. maltophila* strains and the β-proteobacterium *Variovorax* sp. CF313 were more abundant in manured soil and *Gemmatimonas aurantiaca* T27, the actinobacteria *Janibacter* sp. HTCC2649, *I. coccineus* YM16304, *Nocardioides* sp. CF8 and *Ca*. Microthrix parvicella, the δ-proteobacterium *Haliangium ochraceum* and the Euryarchaeoton *Ca*. Halobonum tyrrellensis were more abundant in inorganically fertilized soil, particularly fertilized^{-P} soil.

The response of *phoX* to the different soil fertilization practices were distinctly different from that of *phoD*. In this case, both gene normalized abundance and $BWPD_1$ (Figures 5A and B) were significantly different across the treatments and lowest in fertilized soil. Normalized abundance in fertilized soil was significantly lower than for the other treatments (smallest difference, $Q = 5.7$; $p =$

0.016). No other significant differences in normalized abundance were detected. Phylogenetic diversity in the fertilized soil was significantly reduced compared to diversity in the fertilized^{-P} soil (Q = 5.9; p = 0.014). No other differences in phylogenetic diversity were significant. Edge-PCA ordination (Figure 5C) based upon phylogenetic placement of metagenome reads (Figure 5D) showed less consistency within treatments than was evident for *phoD*. The primary axis separated fertilized soil from fertilized^{-P} soil: fertilized^{-N} and manured soils were intermediate between these two extremes. Ecotypes homologous to *phoX* genes of the β-proteobacteria *Variovorax paradoxus* B4, *Acidovorax* sp. CF316, and *Candidatus* Accumulibacter phosphatis had increased abundance in fertilized soil while ecotypes with homology to genes of the β-proteobacteria *Limnobacter* sp. MED105, Burkholderiales bacterium JOSHI001, *Janthinobacterium lividum*, *Ramlibacter tatouinensis* TTB310 and *Leptothrix cholodnii* SP6, and the γ-proteobacterium *Thioalkalimicrobium aerophilum* AL3 were less abundant in fertilized soil and more abundant in fertilized^{-P} soil. The second axis essentially separated manured soil from those receiving inorganic fertilization. Ecotypes associated with differences on this axis had homology to *phoX* genes of *Ca*. Accumulibacter phosphatis, *Acidvorax* sp. CF316 and NO1, *V. paradoxus* B4, *T. aerophilum* AL3 and *Sphingomonas* sp. S17 which were all more abundant in soil receving inorganic fertilizer, and ecotypes with homology to the γ-proteobacteria HdN1, *Pseudomonas* sp. BAY1663 and *Alcanivorax dieselolei* B5 which were more abudant in manured soil.

For the *phoA* gene, significant treatment effects were evident for both normalized abundance and $BWPD_1$ (Supplementary Figure 1). Abundance was significantly greater in fertilized soil than any other treatment (smallest difference, Q = 5.8, p = 0.015) and fertilized soil also presented significantly lower $BWPD_1$ than the other treatment (smallest difference, Q = 7.0, p = 0.005). Edge-PCA ordination showed limited clustering according to treatment. Fertilized soil was separated from the other treatment soils on the primary axis (accounting for 65% of the variability in abundance-weighted placements) largely as a result of increased abudance of ecotypes with homology to *phoA* sequences of the actinobacteria *Streptoalloteichus hindustanus*, *Amycolatopsis azurea* and *A. mediterrani*. There was little separation of treatments on the second axis.

Non-specific acid phosphatase genes – Of the three classes of NSAPs, class B was not found in any substantial numbers (less than 16 reads per metagenome)

and so was not analysed further. However, both classes A and C were found in significant numbers and responded to soil treatment. Significant effects of soil treatment were evident for class A gene normalized abundance and phylogenetic diversity (Figures 6A and B). Class A normalized abundance in fertilized soil was significantly greater than in either fertilized^{-N} or fertilized^{-P} soils (smallest difference, $Q = 8.2$, $p = 0.002$): these two treatments were associated with the lowest normalized abundance of all treatments. No other abundance comparisons were significant. In addition, class A gene $BWPD_1$ was also greatest in fertilized soil and significantly greater than ecotype $BWPD_1$ in either manured or fertilized^{-N} soils (smallest difference, $Q = 5.4$, $p = 0.022$). Fertilized^{-P} soils were associated with intermediate $BWPD_1$ and not significantly different from either group of treatments. Edge-PCA (Figure 6C) of phylogenetic placement of metagenome reads (Figure 6D) separated fertilized soils from the other treatments on the principal axis which accounted for over 80% of the variability. Manured soil was separated from fertilized soils on the second axis which accounted for only 8% of the variability. On axis 1, ecotypes which were most abundant in fertilized soil showed homology to class A genes from the closely-related α-proteobacteria *Sphingomonas parapaucimobilis*, two genes associated with *Phenylobacterium* sp. Root700, *Phenylobacterium* sp. Root77, *Phenylobacterium zucineum* HLK1. Ecotypes showing homology to class A genes of the γ-proteobacteria *Lysobacter dokdonensis* DS58 and *Methylococcus capsulatus*, the cyanobacterium *Synechococcus* sp. RS9916, the δ-proteobacterium *Desulfotalea psychrophila* LSv54, and the β-proteobacterium *Variovorax* spp. were more abundant in manured, fertilized^{-N} and fertilized^{-P} soil. On the second axis, ecotypes with homology to class A genes of the α-proteobacteria *Novosphingobium* sp. AP12 and *Caulobacter crescentus* OR37, *D. psychrophila* LSv54 and *Variovorax* sp. Root437 were more abundant in manured soil, while ecotypes with homology to genes of *Novosphingobium tardaugens,* and the α-proteobacteria *Brevundimonas* spp. AAP58 and Leaf363 were more abundant in fertilized^{-N} and fertilized^{-P} soils.

For class C genes, the was no significant effect of soil treatment upon gene normalized abundance (Figure 7A). However, as with class A genes, $BWPD_1$ was greatest in fertilized soil (Figure 7B), significantly greater than in manured or fertilized^{-N} soil (smallest difference, $Q = 4.6$, $p = 0.048$). Manured soil presented the least diverse assemblage of NSAPc ecotypes of all the soils. No other diversity comparisons were significantly different. Accordingly, edge-PCA (Figure 7C) of phylogenetic placement of metagenome reads (Figure 7D) separated manured soil

from all fertilized soils on the primary axis in a pattern consistent with 16S rRNA gene distribution. Manured and fertilized treatments formed defined clusters and separated on this axis, but fertilized^{-P} and fertilized^{-N} treatments formed less distinct clusters, separated to some degree on the secondary axis. Ecotypes most responsible for separation on edge-PCA axis 1 had homology to NSAPc genes of *Nitrospira defluvii*, the γ-proteobacterium SG8-31 and the α-proteobacterium BAL199 which all showed increased abundance in manured soil. In fertilized soil, ecotypes with homology to NSAPc genes associated with the Bacteroidetes *Solitalea canadensis*, *Flavobacterium* sp. Leaf82, *Chitinophaga pinensis* and *Mucilaginibacter paludis*, the δ-proteobacteria *Bacteriovorax* sp. and *Halobacteriovorax marinus* and the γ-proteobacterium *Lysobacter dokdonenesis* DS58 were more abundant. On axis 2, ecotypes with homology to genes of *N. defluvii*, Nitrospira bacterium SG8-3, the γ-proteobacteria *Cellvibrio* sp. and SG8-31, *Caldithrix abyssi*, the Bacteroidiales bacterium TBC1 and a number of *Lysobacter* spp. all increased in abundance in fertilized^{-N} and fertilized^{-P} soils. Ecotypes associated with increased abundance in fertilized soils on axis 2 showed homology to genes of *M. paludis,* the α-proteobacteria BAL199 and *Iniquilinus limosus* MP06 and the γ-proteobacterium *Arenimonas oryziterrae.*

Myo-inositol hexakisphosphate phosphohydrolase genes – Of the three phytase genes studied here, βPPhy genes were most abundant, although less abundant than phosphatase genes. The βPPhy was unusual amongst the genes in that it showed greatest abundance in the manured and fertilized^{-P} soils (Figure 8A). Normalized abundance in these two soils was significantly greater than in either fertilized or fertilized^{-N} soils (smallest difference, $Q = 18.0$ $p < 0.001$). The same pattern was observed for $BWPD_1$ (Figure 8B), where manured and fertilized^{-P} soils were associated with more phylogenetically diverse assemblages than either the fertilized or fertilized^{-N} soils. However, only the difference in diversity between manured and fertilized soils was significant ($Q = 5.7$, $p = 0.016$). Ordination of the treatments using edge-PCA (Figure 8C) based upon the phylogenetic placement of metagenome reads (Figure 8D) supported this observation: clusters were not well defined according to treatment, except that the fertilized soil was separated from the other soils on the primary axis. The dominant ecotype in all soils showed high homology to genes associated with the cyanobacterium *Cyanothece,* however changes in the abudance of these ecotypes were not associated with the observed separation of treatments by edge-PCA. The first edge-PCA axis separated manured from fertilized soil in a manner similar to the

16S rRNA gene. Ecotypes responsible for this separation displayed homology to genes of the firmicute *Paenibacillus* spp. and the γ-proteobacterium *Pseudomonas stutzeri,* both least abundant in fertilized soil. Ecotypes with homology to genes associated with *Pseudomonas fluorescens* F113 and a clade of *Pseudoalteromonas* spp. were more abundant in fertilized soil but were present at very low abundance. The second edge-PCA axis effectively separated fertilized soil from fertilized^{-N} and fertilized^{-P} soils. Ecotypes with homology to genes of *Paenibacillus* and *Pseudomonas stuzeri* were more abundant in N- and P- lacking soils and ecotypes with homology to βPPhy genes associated with *Pseudomonas fluorescens*, *Pseudoaltermonas* and *Janthinobacterium* were more abundant in fertilized soil although as stated abive these were at veryu low abundance.

For the acidic phytases there were only limited differences between treatments (Supplementary Figures 2 and 3). In both cases, there was no significant effect of soil management upon gene $BWPD_1$ and normalized abundance was significantly greater in fertilized soil than either fertilized^{-P} or fertilized^{-N} soils (CPhy smallest difference, $Q = 6.5$, $p = 0.009$; HAP smallest difference, $Q = 5.7$, $p = 0.016$). The causes of this were essentially that both genes were dominated by a limited number of ecotypes. In the case of CPhy, the most abundant ecotypes in any soil showed homology to the CPhy gene of the δ-proteobacterium *Bdellovibrio bacteriovorus* and the limited separation of treatments evident from edge-PCA was largely due to differences in the abundance of these ecotypes, which were particularly high for one fertilized pseudo-replicate. For HAP genes, the dominant ecotypes showed homology to a broad range of HAP genes associated with enteric and pathogenic bacteria including the γ-proteobacteria *Escherichia*, *Yersinia*, *Edwardsiella, Serratia* and *Plesiomonas shigelloides* and the β-proteobacteria *Burkholderia* sp. and *Pandoreae* sp. Again, there was limited separation of treatments by edge-PCA.

Edaphic factor influence on gene distribution – Significant differences were observed in the distribution of the combined genes in response to the experimental treatments (PERMANOVA, *pseudo-F* = 2.6, $p < 0.001$). The effect of differences in edaphic factors resulting from long-term management upon the distribution of phosphohydrolase genes distribution was determined by constrained correspondence analysis (CCA, Figure 9). Forward selection identified six of the eleven parameters (shown in Table I) as exerting a significant influence. These were pH (*pseudo-F* = 3.2, $p = 0.040$), K_{ex} (*pseudo-F* = 4.1, $p = 0.005$), C/N

(*pseudo-F* = 3.0, $p = 0.042$), grain yield (*pseudo-F* = 2.9, $p = 0.032$), Ca_{ex} (*pseudo-F* = 2.7, $p = 0.013$) and %N (*pseudo-F* = 2.4, $p = 0.042$). Noteably, neither Olsen-P or P-balance were associated with significant influence upon phosphohydrolase gene distribution. The CCA model accounted for a significant proportion of variability and both axes of the resulting ordination accounted for significant proportions of the variability explained by the model. Axis 1 (accounting for 60% of variability accounted for by the model) seperated fertilized soil from the other three, grouping manured soil with fertilized^{-N} and fertilized^{-P} soils. The edaphic factors most clearly associated with axis 1 were Ca_{ex} and pH, both high in the manured, fertilized^{-N} and fertilized^{-P} soils and reduced in fertilized soil. Axis 2 (accounting for 14% of variability) separated manured soil from all inorganically-fertilized soils. The edaphic factors most clearly associated with this separation were total N (%N) and the C/N ratio, both greatest in manured soil and lowest in inorganically-fertilized soils. The remaining edaphic factors were less closely associated with Axis 2 (K_{ex} and grain yield). Examination of the distribution of the eight individual gene families in the context of the edaphic factors measured across the expierment, suggest separation both between the larger groupings of alkaline phosphatase, acid phosphatase and phytase families, and between individual genes. Within these groups, the alkaline phosphatase *phoX* and the β-propeller phytase appear to have rather restricted distribution within the experiment parameter space than the other genes. The most abundant gene, the alkaline phosphatse *phoD*, dominated experimental parameter space with less abundant genes distributed on the periphery of the *phoD* distribution. Much of the individual gene distribution was associated with the range of Ca_{ex}, but both NSAP genes appeared also to be responsive to reduced total N and reduced C/N ratios and HAP and CPhy genes appear to be associated with reduced pH. Grain yield, used here as a proxy for competition with wheat plants for P, did not exert any evident influence upon phosphohydrolase gene distribution.

Discussion

There is a great deal of evidence that soil management exerts a direct influence upon soil microbial communities: both acute and chronic stresses can lead to potentially irreversible changes in community structure[20]. Community changes are often obvserved in agricultural soils and between different agricultural practices. For example, soils managed organically, that is by the addition of animal-derived or green manures, typically contain more abundant and active microbial communities than inorganically-fertilized soils[21]. These communities are

often also more diverse[22-25]. On the Broadbalk winter wheat experiment, addition of farmyard manure or inorganic fertilizer for over 150 years has not induced significant changes to 16S rRNA phylogenetic diversity. This difference may arise for a number of technical and experimental reasons. One explanation is that the diversity measure used here, phylogenetic diversity, differs from the more traditional non-parametric measures used to describe diversity elsewhere[22,26] and incorporates information regarding gene phylogenetic structure. In this sense, phylogenetic diversity effectively provides a proxy for *feature* or *attribute* diversity[27] presenting a more biologically-meaningful assessment of assemblage diversity. Additionally, 16S rRNA phylogenetic diversity is assessed here based upon phylogenetic placement of shotgun metagenome reads rather than amplicon-based assessment[22,24-26] and may present reduced bias in the assessement, since it does not rely upon primer-based amplification. The unprecedented timespan of the Broadbalk experiment may also explain the lack of differences in phylogenetic diversity, in the sense that communities subject to the different treatments have effectively had time reach stable end-point structures, less likely in shorter experiments. Although phylogenetic diversity was consistent between treatments, disctinct 16S rRNA assemblages were detected under the different treatments and directly influenced by the nature of the nutrient amendement, although the specific response of these assemblges differed from those observed on the Swiss DOK long-term experiment[28]. On Broadbalk, organisms such as *Pirellula staleyi*, *Woeseia oceani* and *Steriodobacter denitrificans* were most abundant in manured soils. Rather unexpectedly, the obligate nematode endosymbiont[29] *Candidatus* Xiphinematobacter was most abundant in, and characteristic of fertilized soil: dagger nematodes (*Xiphinema* spp.) are ectoparasites of various plant crop species, including cereals, and potentially act as vectors for several economically important plant viruses[30].

Estimates of mean phylogenetic diversity indicate that the assemblages of phosphohydrolase gene ecotypes are more sensitive to fertilization than the 16S rRNA-conditional microbial community and respond in a fundamentally different manner. Despite sharing a common function, each phosphohydrolase gene has a distinct abundance and phylogenetic diversity profile – and response to management - within the soil communities studied here. The alkaline phosphatase *phoD* was the most abundant and phylogenetically diverse phosphohydrolase gene. This gene is the most abundant alkaline phosphatase in both marine and soil systems[14, 31], but it is clear that *phoD* is also more abundant

than any non-specific acid phosphatse and phytase genes. A group of genes appear to share a similar abundance, but different phylogenetic diversity: these include a second alkaline phosphatase, *phoX*, class A and C NSAPs and βPPhy genes. Of these, *phoX* is the most abundant and phylogenetically diverse and βPPhy the least abundant and diverse. The remaining genes are all of low abundance and phylogenetic diversity, suggesting that alkaline phosphatase *phoA*, class B NSAP and the acidic phytase CPhy and HAPhy genes do not contribute significantly to hydrolysis of P_{org} in soils, consistent with their presence largely in enteric or pathogenic bacteria[14,16]. Based upon the substrate specificity of the different enzyme groups[31,32] the observed abundance of phosphohydrolase genes in soil suggests that phosphomono- and phosphodiesters are the principal sources of P for soil microbes, and that phytate is not a major source of P.

The observation that neither P-associated parameter (bioavailable P as Olsen-P and P-balance describing the likely competition for P with crop plants) within the soils was significantly associated with the distribution of phosphohydrolase genes was unexpected and contrary to previous studies of the effects of edaphic factors upon gene diversity[33], although not without precedent[34]. It is noteworthy that observations of the effect of P chemistry upon gene diversity were apparent from a study of soils from a wide variety of geographical locations, and with a wide variety of different P-chemistry and availability[33], whereas this study and that of soil from the Glenlea Research Station, Manitoba[35] are local studies, comparing soils of relatively similar chemistry *etc*. It is possible that different edaphic factors influence gene distribution and diversity at global and local scales.

The detailed analysis of gene abundance presented here reveals the *realized ecological niche*[35] of each gene within the multidimensional space described by soil edaphic factors, most clearly demonstrasted by CCA (Figure 9). The distribution of phosphohydrolase genes in Broadbalk soils suggest that pH and Ca_{ex} exert a significant, dominant influence. Soil pH has long been known to affect the diversity of soil microbial communities[3] and for phosphatases and phytases, the pH optima - at least *in vitro* - are well established: pH 8 for alkaline phosphatases PhoD and PhoX[36,37], 9 – 6 for βPPhy[38], 7.5 – 5 for NSAPs[32] and 5 – 4 for acidic phytases[39]. The realized ecological niches evident in Broadbalk soil are consistent with these relative optima, with the alkaline phosphatses relatively more abundant at high pH and the acid phytases relatively more abundant at lower pH. Niche sensitivity to Ca_{ex} may arise as a consequence of both abiotic and biotic interactions. In some calcareous soils, the organic proportion of total P is correlated positively with Ca content[9] and thus the relatively higher abundance of phosphatases at raised Ca_{ex} concentrations could reflect this. On Broadbalk, Ca is derived from liming of soil to maintain optimal pH for crop yield. Alternatively, PhoD and PhoX both require Fe^{3+} and Ca^{2+} as cofactors[36,37] and BPPhy also requires Ca^{2+} [40]. The other proteins all require different co-factors; Zn^{2+} and Mg^{2+} for PhoA[41], VO_4^{3-} for the NSAPs[42]. The acid phytases CPhy and HAPhy have no known requirement for metal co-factors. Thus, the increased abudance of *phoD*, *phoX* and βPPhy genes at raised Ca_{ex} concentrations (and reduced abudance at lower Ca_{ex} concentrations) is consistent with their cofactor requirements. If this is the case, it is evident that some ecotypes, for example those with homology to the *Rhodanobacter spathiphylli* B39 *phoD* gene and the *Ca*. Accumulibacter phosphatis *phoX* gene appear to code for enzymes which are more efficient at reduced Ca^{2+} availability since they are more abundant in fertilized soil with reduced Ca_{ex} concentrations. In contrast, the general distribution of *phoX* and βPPhy ecotypes is much more restricted, suggesting greater sensitivity to Ca concentrations. Whatever the cause, the result is that function - in this case hydrolysis of P_{org} compounds to release orthophosphate – is maintained within the background of changing community structure. However, the genes (and thus proteins) responsible for this function are not consistent and

are selected based upon environmental fitness, probably to chemical edaphic

factors, but not P-availability or plant growth.

Genes coding for non-essential accessory proteins, typically associated with central intermediary metabolism[43], undergo frequent horizontal transfer driven by environmental factors such as nutrient availability and growth conditions[19]. These genes represent a supraspecific gene pool[35] that can increase the fitness of individuals of very different species. In a community, organisms from many different species may have access to this supraspecific gene pool, composed of transferable genes present in the community's individual cells and associated mobile genetic elements, or introduced by arrivals from other communities. However, not all potential prokaryotic genomes within a community may represent a suitable venue for an incoming gene to persist. To be maintained in a stable fashion for any significant time, a gene would need to encode a function that conferred a valuable adaptation to the organism in its current environment. The data presented here relating to phosphohydrolase genes suggests that pH and Ca_{ex} both contribute to this selectivity. An incoming gene that does not provide a novel or superior capability to that already existing in a recipient genome would not be maintained, would not spread through the recipient population and could potentially be lost due to the high substitution and deletion rates operating in prokaryotic genomes[44].

Materials and Methods

Field Experiment and Sampling - Soil was sampled from four treatments of the Broadbalk Long-Term field experiment (Macdonald *et al.*, 2018) (51˚48’35” N, 00˚22’30” W). The experimental soil is a clay loam to silty clay loam over clay-with flints (FAO Classification: Chromic Luvisol). The soil contains many flints and is slightly calcareous. The experiment is under-drained and so free draining. The plough layer (0-23 cm) is limed when necessary to maintain a minimum soil pH of 7.0 – 7.5. The four treatment plots are on Section 1 of the experiment, which has been sown continuously with winter wheat (*Triticum aestivum* L., most recently Crusoe seed coated with Redigo® Deter® combination insecticide/fungicide treatment, Bayer CropScience) since 1843, except for ocassional fallow years to control weeds. We compared plots to which farmyard manure (from cattle) has been added at a rate of 35 t ha^{-1} per year since 1843 (manured), with plots to which inorganic fertilizer containing 144 kg ha^{-1} nitrogen (N) since 1986 as Nitram® ammonium nitrate, 35 kg ha^{-1} phosphorus (P) as triple superphosphate (calcium dihydrogen phosphate), 90 kg ha^{-1} potassium (K) as potassium sulfate, and 12 kg ha^{-1} magnesium (Mg) as kieserite per year (fertilized) has been added since 1852, inorganic fertilization lacking N-fertilization but receiving 35 kg ha^{-1} P, 90 kg ha^{-1} K and 12 kg ha^{-1} Mg per year (fertilized^{-N}) since 1852 and inorganic fertilization lacking P-fertilization but receiving 192 kg ha^{-1} N, (96 kg ha^{-1} N 1906-2000) 90 kg ha^{-1} K and 12 kg ha^{-1} Mg per year since 1906 (fertilized^{-P}). Current plot size is 28 x 6m, except plot 2.2 which is 28 x 4m. Since 2001, fertilized and fertilized^{-N} plots have not received P as it was considered in excess. The soil is tilled conventionally. All treatments were sampled in October 2015. Since treatments are not replicated on the field experiment, three *pseudo*-replicates were collected from each treatment plot. All sampling equipment was cleaned with 70% ethanol between samples. The top 10 cm of soil was sampled with a 3-cm diameter auger. For each *pseudo*-replicate, ten cores were pooled and thoroughly mixed whilst sieving through a 2-mm mesh. Samples were then frozen and stored at -80 °C.

Chemical properties of Broadbalk soils and crop performance have been measured routinely since the experiment inception. Data were taken from the *e*-RA database[45] maintained by Rothamsted Research. Plant-available (Olsen) phosphorus (P) was extracted in 0.5M sodium bicarbonate before being measured most recently on a NexION® 300X inductively-coupled plasma mass spectrometer (Perkin Elmer LAS (UK) Ltd., Seer Green, UK). Exchangeable potassium (K_{ex}), calcium (Ca_{ex}), magnesium (Mg_{ex}) and sodium (Na_{ex}) concentrations were measured on an Optima® inductively-coupled plasma spectrometer (ICP-OES, Perkin Elmer) following extraction in a 1M ammonium acetate solution (pH 7). Total nitrogen (N) was measured by combustion using a Leco® TruMac® analyser (LECO (UK), Stockport, UK) and soil organic carbon (SOC) was measured by ultra-violet oxidation using a TOC-V WP Analyzer (Shimadzu UK Ltd., Milton Keynes, UK). Soil pH was measured in water (1:2.5 soil: solution). Grain yield for each plot was measured at 85% dry matter. A P-balance was calculated for each soil based upon the difference between P-inputs in manure or fertilizer and P removed at harvest in grain and straw (measured by ICP-OES of digested plant material). This calculation did not account for P in chaff or stubble left in the field. Each parameter is not measured every year. Since the structure of the soil microbiomes will reflect the long-term soil chemistry in each treatment, the three most recent measurements for each parameter were used to calculate mean estimates (2000, 2005 and 2010). All estimates were generated since P-addition to fertilized and fertilized^{-N} treatments were suspended.

DNA Extraction, Sequencing and Quality Control - Soil community DNA was extracted from a minimum of 2 g of thawed soil using MoBio PowerSoil® DNA isolation kits (Mo Bio Laboratories, Inc. Carlsbad, CA). DNA quantification and quality control was assessed using a Qubit 2.0 fluorimeter (Thermo Fisher Scientific, Waltham, USA) and 2100 Bioanalyzer DNA chips (Agilent Technologies, Santa Clara, USA). 10 µg of high-quality DNA was provided for sequencing for each of the twelve samples. Shotgun metagenomic sequencing of DNA was performed using 150 base paired-end chemistry on an Illumina® HiSeq™ 2500 sequencing platform by Beijing Novogene Bioinformatics Technology Co. Ltd. (Beijing, China). The generated raw sequences were limited to a minimum quality score of 25 and a minimum read length of 70 bases using Trimmomatic[46]. After filtering to remove substandard sequences, the average number of metagenome reads for each soil was $3.44x10^8$ for manured soil, $3.69x10^8$ for fertilized soil, $3.08x10^8$ for fertilized^{-P} soil, and $3.76x10^8$ for fertilized^{-N} soil (range across all datasets $2.95x10^8$ – $4.08x10^8$ reads).

Estimation of gene abundance and phylogeny - Each of the twelve metagenomes generated in this study were analysed to estimate the abundance of the 16S rRNA gene and each of nine phosphohydrolase genes. Nucleotide-based profile hidden Markov models (pHMM) were generated from multi-sequence alignments (MSAs) of reference sequences of each gene using hmmbuild, part of the HMMER ver 3.1 suite[47]. MSAs were generated using the *E-INS-i* iterative refinement algorithm in MAFFT version 7.3[48] using the 1PAM/κ=2 scoring matrix. For the 16S rRNA gene, the pHMM was generated using the set of 4,528 reference sequences associated with paprica[49], built December 2017. For phosphohydrolase genes, pHMMs were generated from reference sequences of the alkaline phosphatase *phoD*, *phoX* and *phoA*, and β-propeller (βPPhy), cysteine (CPhy) and histidine acid (HAPhy) phytase described by Neal *et al*.[14], and for classes A, B and C of non-specific acid phosphatase (NSAP) described by Neal *et al*.[16]. Metagenome reads with homology to the pHMMs were identified using hmmsearch with a $1x10^{-5}$ Expect-value cut-off. To compare abundance of each gene relative to the 16S rRNA read count between metagenomes, a length-normalized abundance was calculated as follows: relative abundance = (PHO read count x PHO pHMM length)*100/(16S read count x 16S pHMM length), where PHO represents one of the nine phosphohydrolase genes, 16S represents the 16S rRNA gene and pHMM length is in nucleotide bases.

Metagenome reads showing homology to each gene were assigned to branches of phylogenetic trees generated from the respective reference gene sets using a phylogenetic placement algorithm, pplacer version 1.1alpha10[50] and visualized using iTOL version 4.2.3[51]. For the 16S rRNA gene, these placements can be translated into robust relative abundance estimates of named organisms using the taxonomic labelling of the tree branches. This is not the case for the phosphohydrolase genes where instead, placement indicates the degree of homology of the metagenome reads (ecotypes) to the respective genes found in sequenced organisms, identified by taxonomic labels of the tree branches.

Statistical Analysis - The effects of different fertilization treatments upon edaphic factors, and estimates of normalized abundance and α-diversity for each gene were analysed using analysis of variance (ANOVA) after testing for homogeneity of variances using Levene's test and normality using the Shapiro-Wilk test. Data for some genes were associated with significantly non-normal distributions, although the variances were homogenous. Where this was the case, distribution-free tests of significance of *F*-values based upon 99,999 permutations were adopted. Where significant treatment effects were identified, *post-hoc* pair-wise comparisons were performed using Tukey-Kramer Studentized *Q*, following the Copenhaver-Holland multiple comparison procedure[52]. All tests were calculated using PAST version 3.2[53]. For all tests, an α of 0.05 was considered significant.

Estimates of gene α-diversity based upon phylogenetic placement of homologous metagenomic reads were assessed by computing the balance-weighted phylogenetic diversity[54] ($BWPD_1$) of each sample using the guppy fpd binary (part of the pplacer package), accounting for pendant branch length (--include-pendant flag). To assess the depth of sequencing of the soil communities compared with the total diversity of the nine genes within them, rarefaction curves of expected mean phylogenetic diversity[55] were generated using the guppy rarefact binary, interpreting placement weights as counts (--weight-as-count flag) and calculating up to a rarefaction size (*k*) of 70,000. Additionally, principal component analysis of the difference in placement densities on reference tree branches, termed edge-PCA[56], was used for graphical representation of phylogeny-based differences between treatments in a two-dimensional plane using the guppy epca binary and treating each query as a point mass concentrated on the highest-weight placement (--point-mass flag). One advantage of edge-PCA is that branches associated with placements contributing to eigenvalues on each axis are identified.

Canonical correspondence analysis (CCA) was employed to determine which measured edaphic factors may contribute to the observed gene distribution. Edpahic factors shown in Table I associated with a significant influence upon gene ecotype distribution were identified by forward selection procedures in CANOCO version 5[57]. These factors were then included in a CCA model in PAST version 3.2 based upon the eigenanalysis algorithm of Legendre & Legendre[58]. The validity of the resulting model and ordination was determined based upon 999 permutations.

Acknowledgements

This work is supported by the BBSRC-funded *Soil to Nutrition* strategic programme (BBS/E/C/000I0310) and jointly by the Natural Environment Research Council and BBSRC as part of the *Achieving Sustainable Agricultural Systems* research programme (NE/N018125/1 LTS-M). The Rothamsted Long-term Experiments National Capability is supported by the Lawes Agricultural Trust and the Biotechnology and Biological Sciences Research Council (Grants BBS/E/C/00005189 (2012-2017) and BBS/E/C/000J0300 (2017-2022)). The authors a grateful to Yuan Liu for assistance with CANOCO, to members of Rothamsted's molecular microbial ecology research group for help with soil sampling, and Andrew MacDonald for comments which helped to improve a draft of the manuscript.

References

1. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD *et al*., (2018). A global atlas of the dominant bacteria found in soil. *Science* 359, 320-325.

2. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM *et al*., (2018). Structure and function of the global topsoil microbiome. *Nature* 560, 233-237.

3. Fierer N, Jackson RB (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the U. S. A*. 103, 626-631.

4. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrumpf M, Daniela R (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Scientific Reports* 6, article number 33696.

5. Haygarth PM, Ritz K (2009). The future of soils and land use in the UK: Soil systems for the provision of land-based ecosystem services. *Land Use Policy* 26, S187-S197.

6. Lehman RM, Cambardella CA, Stott DE, Acosta-Martinez V, Manter DK, Buyer JS, Maul JE, Smith JL, Collins HP, Halvorson JJ, Kremer RJ, Lundgren JG, Ducey TF, Jin VL, Karlen DL (2015). Understanding and enhancing soil biological health: the solution for reversing soil degradation. *Sustainability* 7, 988-1027.

7. Crowther TW, Todd-Brown KEO, Rowe CW, Wieder WR, Carey JC, Machmuller MB, Snoek BL, Fang S, Zhou G, Allison SD, et al., (2016). Quantifying global soil carbon losses in response to warming. *Nature* 540, 104-108.

8. Kallenbach CM, Frey SD, Grandy AS (2016). Direct evidence for microbial-derived soil organic metter formation and its ecophysiological controls. *Nature Communications* 17: 13630

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

10. Richardson AE, Simpson RJ (2011). Soil microorganisms mediating phosphorus availability. *Plant Physiology* 156, 989-996.

11. Haygarth PM, Harrison AF, Turner BL (2018). On the history and future of soil organic phosphorus research: a critique across three generations. *European Journal of Soil Science* 69, 86-94.

12. Sebastian M, Ammerman JW (2009). The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME Journal* 3, 563-572.

13. Zaheer R, Morton R, Proudfoot M, Yakunin A, Finan TM (2009). Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria. *Environmental Microbiology* 11, 1572-1587.

14. Neal AL, Rossmann M, Brearley C, Akkari E, Guyomar C, Clark IM, Allen E, Hirsch PR (2017). Land-use influences phosphatase gene microdiversity in soils. *Environmental Microbiology* 19, 2740-2753.

15. Gandhi NU, Chandra SB (2012). A comparative analysis of three classes of bacterial non-specific acid phosphatases and archael phosphoesterases: evolutionary perspective. *Acta Informatica Medica* 20, 167-173.

16. Neal AL, Blackwell M, Akkari E, Guyomar C, Clark I, Hirsch PR (2017b). Phylogenetic distribution, biogeography and the effects of land management upon bacterial non-specific acid phosphatase gene diversity and abundance. *Plant and Soil* 427, 175-189.

17. Lim BL, Yeung P, Cheng C, Hill JE (2007). Distribution and diversity of phytate-mineralizing bacteria. *ISME Journal* 1, 321-330.

18. Pál C, Papp B, Lercher MJ (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics* 12, 1372-1375.

19. Takemoto K (2012). Current understanding of the formation and adaptation of metabolic systems based on network theory. *Metabolites* 2, 429-457.

20. Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H *et al*., (2012). Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology* 3, article 417.

21. Lori M, Symnaczik S, Mäder P, de Deyn G, Gattinger A (2017). Organic farming enhances soil microbial abudance and activity – a meta-analysis and meta-regression. *PLoS One* 12(7), e0180442.

22. Zhen Z, Liu H, Wang N, Guo L, Meng J, Ding N, Wu G, Jiang G (2014). Effects of manure compost application on soil microbial community diversity and soil microenvironments in a temperate cropland in China. *PLoS One* 9, e108555.

23. Francioli D, Schulz E, Lentendu G, Wubet T, Buscot F, Reitz T (2016). Mineral vs. organic amendements: microbial community structure, activity and abundance of agriculturally relevant microbes are driven by long-term fertilization strategies. *Frontiers in Microbiology* 7, 1446.

24. Lupatini M, Korthals GW, de Hollander M, Janssens TKS, Kuramae EE (2017). Soil microbiome is more heterogeneous in organic than in conventional farming systems. *Frontiers in Microbiology* 7, 2064.

25. Li F, Chen L, Zhang J, Yin J, Huang S (2017). Bacterial community structure after long-term organic and inorganic fertilization reveals important associations between soil nutrients and specific taxa involved in nutrient transformations. *Frontiers in Microbiology* 8, 187.

26. Xun W, Zhao J, Xue C, Zhang G, Ran W, Wang B, Shen Q, Zhang R (2016). Significant alteration of soil bacterial communities and organic carbon decomposition by different long-term fertilization management conditions of

extremely low-productivity arable soil in South China. *Environmental Microbiology* 18, 1907-1917.

27. Chao A, Chiu C-H, Hsieh TC, Davis T, Nipperess DA, Faith DP (2014). Rarefaction and extrapolation of phylogenetic diversity. *Methods in Ecology and Evolution* 6, 380-388.

28. Hartmann M, Frey B, Mayer J, Mäder P, Widmer F (2015). Distinct soil microbial diversity under long-term organic and conventional farming. *ISME Journal* 9, 1177-1194.

29. Vandekerckhove TTM, Coomans A, Cornelis K, Baert P, Gillis M (2002). Use of the *Verrucomicrobia*-specific probe EUB338-III and flurescent in situ hybridization for detection of "*Candidatus* Xiphinematobacter" cells in nematode hosts. *Applied and Environmental Microbiology* 68, 3121-3125.

30. McFarlane SA, Neilson R, Brown DJF (2002). Nematodes. In: Plumb RT (Ed.) *Advances in Botanical Research*. Academic Press, San Diego, CA, USA, 169–198. doi: 10.1016/s0065-2296(02)36063-4

31. Luo H, Benner R, Long RA, Hu J (2009). Subcellular localization of marine bacterial alkaline phosphatases. *Proceedings of the National Academy of Sciences of the U. S. A*. 106, 21219-21223.

32. Rossolini GM, Schippa S, Riccio ML, Berlutti F, Macaskie LE, Thaller MC (1998). Bacterial nonspecific acid phosphatases: physiology, evolution and use as tools in microbial biotechnology. *Cell and Molecular Life Sciences* 54, 833-850.

33. Ragot SA, Kertesz MA, Mészáros É, Frossard E, Bünemann EK (2017). Soil *phoD* and *phoX* alkaline phosphatase gene diversity responds to multiple environmental factors. *FEMS Microbiology Ecology* 93, fiw212

34. Fraser T, Lynch DH, Entz MH, Dunfield KE (2015). Linking alkaline phosphatase activity with bacterial *phoD* gene abudance in soil from a long-term management trial. *Geoderma* 257-258, 155-122.

35. Francino MP (2012). The ecology of bacterial genes and the survival of the new. *International Journal of Evolutionary Biology* 2012, article ID 394026.

36. Rodriguez F, Lillington J, Johnston S, Timmel CR, Lea SM, Berks BC (2014). Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. *Journal of Biological Chemistry* 289, 30889-30899.

37. Yong SC, Roversi P, Lillington J, Rodriguez F, Krehenbrink M, Zeldin OB, Garman EF, Lea SM, Berks BC (2014). A complex iron-calcium cofactor catalysing phosphotransfer chemistry. *Science* 345, 1170-1173.

38. Zhang R, Yang P, Huang H, Yuan T, Shi P, Meng K, Yao B (2011). Molecular and biochemical characterization of a new alkaline β-propeller phytase from the insect symbiotic bacterium *Janthinobacterium* sp. TN115. *Applied Microbiology and Biotechnology* 92, 317-325.

39. Huang H, Zhang R, Fu D, Luo J, Li Z, Luo H, Shi P, Yang P, Diao Q, Yao B (2010). Diversity, abudance and characterization of ruminal cysteine phytases suggest their importance in phytate degradation. *Environmental Microbiology* 13, 747-757.

40. Mullaney EJ, Ullah AHJ (2003). The term phytase comprises several different classes of enzymes. *Biochemical and Biophysical Research Communications* 312, 179-184.

41. Torriani, A. (1990). From cell membrane to nucleotides: the phosphate regulon in *Escherichia coli*. *Bioessays* 12, 371–376

42. Littelchild J, Garcia-Rodriguez E, Dalby A, Isupov M (2002). Structural and functional comparisons between vanadium haloperoxidase and acid peroxidase enzymes. *Journal of Molecular Recognition* 15, 291-296.

43. Beiko RG, Harlow TJ, Ragan MA (2005). Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences of the U. S. A*. 102, 14332-14337.

44. Mira A, Ochman H, Moran NA (2001). Deletional bias and the evolution of bacterial genomes. *Trends in Genetics* 17, 589-596.

45. Perryman SAM, Castells-Brooke NID, Glendining MJ, Goulding KWT, Hawkesford MJ, Macdonald AJ, Ostler RJ, Poulton PR, Rawlings CJ, Scott T, Verrier PJ (2018) The electronic Rothamsted Archive (e-RA), an online resource for data from the Rothamsted long-term experiments. *Scientific Data*, **5**, 180072 DOI: 10.1038/sdata.2018.72

46. Bolger AM, Lohse M, Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

47. Eddy SR (2009). A new generation of homology search tools based on probabilistic inference. *Genome Informatics* 23, 205-211.

48. Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30, 772-780.

49. Bowman JS, Ducklow HW (2015). Microbial communities can be described by metabolic structure: a general framework and application to a seasonally variable, depth-stratified microbial community from the Coastal West Antarctic Peninsula. *PLoS ONE* 10.8: e0135868.

50. Matsen FA, Evans SN (2013). Edge principal components and squash clustering: using the spectral signature of phylogenetic placement data for sample comparison. *PLoS One* 8(6), 10.1371/annotation/40cb3123-845a-43e7-b4c0-9fb00b6e2212.

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

52. Copenhaver MD, Holland, B (1988). Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. *Journal of Statistical Computation and Simulation* 30, 1-15.

53. Hammer Ø, Harper DAT, Ryan PD (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1): 9pp.

54. McCoy CO, Matsen FA (2013). Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. *Peer Journal* 1, e157.

55. Nipperess DA, Matsen FA (2013). The mean and variance of phylogenetic diversity under rarefaction. *Methods in Ecology and Evolution* 4, 566-572.

56. Matsen FA, Evans SN (2013). Edge principal components and squash clustering: using the spectral signature of phylogenetic placement data for sample comparison. *PLoS One* 8(6), 10.1371/annotation/40cb3123-845a-43e7-b4c0-9fb00b6e2212.

57. ter Braak CJF (1988). CANOCO - an extension of DECORANA to analyze species-environment relationships. *Vegetation* 75, 159–160.

58. Legendre P, Legendre L (1998). *Numerical Ecology*, 2nd English edition. Elsevier, 853 pp.

FIGURE 1. Phosphorus in soils of the Broadbalk winter wheat long-term experiment. **A** – sodium bicarbonate extractable (Olsen) phosphorus in soils historically receiving farmyard manure, inorganic fertilizer (NPKMg, fertilizer), inorganic fertilizer with no nitrogen addition (fertilizer^{-N}) and inorganic fertilizer with no phosphorus addition (fertilizer^{-P}). Dashed line indicates the estimated Olsen-P of the soil prior to establishment of the experiment in 1856. Addition of phosphorus was halted in 2000 to fertilized and fertilized^{-N} soils. **B** – Gross differences in phosphorus inputs in either manure or fertilizer and take-off associated with wheat grain and straw between the four differently managed soils. Dashed line represents a zero balance of inputs and take-offs.

FIGURE 2. **Effects of soil fertility management upon the microbial assemblages in soil according to 16S rRNA homology**. **A** – estimated mean unrooted phylogenetic diversity (PD – expressed as the sum of branch lengths occupied on the gene phylogenetic tree) of the 16S rRNA gene assemblages at increasing rarefaction size (k) identified in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **B** – box plots of balance-weighted phylogenetic diversity ($BWPD_1$) of 16S rRNA

assemblages in the treatment soils: the median (heavy line), mean (light line) and 5th and 95th percentiles are shown. **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the 16S rRNA gene in each soil on the reference 16S rRNA maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 3. **Phylogenetic Diversity of phosphohydrolase genes in soil**. Estimated mean unrooted phylogenetic diversity (PD – expressed as the sum of branch lengths occupied on reference gene phylogenetic trees) of the alkaline phosphatase (*phoD*, *phoX* and *phoA*), class A (NSAPa) and C (NSAPc) non-specific acid phosphatase and β-propeller (βPPhy), cysteine (CPhy) and histidine acid (HAP) phytase gene assemblages at increasing rarefaction size (*k*) identified in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}).

FIGURE 4. **Effects of soil fertility management upon alkaline phosphatase *phoD* ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the *phoD* gene in each soil on the reference *phoD* maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 5. **Effects of soil fertility management upon alkaline phosphatase *phoX* ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the *phoX* gene in each soil on the reference *phoX* maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 6. **Effects of soil fertility management upon Class A non-specific acid phosphatase (NSAPa) ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the NSAPa gene in each soil on the reference NSAPa maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 7. **Effects of soil fertility management upon Class C non-specific acid phosphatase (NSAPc) ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods

section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the NSAPc gene in each soil on the reference NSAPc maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 8. **Effects of soil fertility management upon β-propeller phytase (βPPhy) ecotype assemblages in soil. A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the βPPhy gene in each soil on the reference βPPhy maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

FIGURE 9. **Apparent phosphohydrolase gene niche separation in soil**. Treatment-conditional triplots based upon constrained correspondence analysis (CCA) of the relationship between phosphohydrolase gene ecotypes and edaphic factors described in Table I. Edaphic factors were selected based upon a forward selection procedure. **A** – ordination indicates that fertilized soil is separated from all other soils on CCA axis 1, sample centroids are represented by triangles and colour-coded according to treatment. Differences in pH and exchangeable calcium (Ca) are most strongly associated with the separation of soils on CCA axis 1. The carbon:nitrogen ratio (C/N), total nitrogen (%N) and exchangeable potassium (K) are most strongly associated with separation of manured from the other soils on CCA axis 2. Values on each axis indicate the percentages of total variation explained by the model explained by the axis, the eigenvalue (λ) and the permutation-based significance of the amount of variation explained by the axis. **B** – **D** same triplot as shown in A, but highlighting association of ecotypes of alkaline phosphatases (B), non-specific acid phosphatases (C) and phytases (D) with the environmental variables, colour coded by gene.

SUPPLEMENTARY FIGURE 1. **Effects of soil fertility management upon alkaline phosphatase *phoA* ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the *phoA* gene in each soil on the reference *phoA* maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

SUPPLEMENTARY FIGURE 2. **Effects of soil fertility management upon cysteine phytase (CPhy) ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic

fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the CPhy gene in each soil on the reference CPhy maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.

SUPPLEMENTARY FIGURE 3. **Effects of soil fertility management upon histidine acid phytase (HAP) ecotype assemblages in soil**. **A** – length-normalized abundance (relative to the 16S rRNA gene, see Materials and Methods section for calculation description) and **B** – balance-weighted phylogenetic diversity ($BWPD_1$) of gene ecotypes in soils receiving farmyard manure, inorganic fertilizer (NPKMg, fertilized), inorganic fertilizer with no nitrogen addition (fertilized^{-N}) and inorganic fertilizer with no phosphorus addition (fertilized^{-P}). **C** – edge-PCA ordination, based upon abundance-weighted phylogenetic placements of metagenome reads with homology to the HAP gene in each soil on the reference HAP maximum-likelihood tree, shown in **D**: shape size is proportional to the number of reads placed at each point on the tree.