Bacterial physiological adaptations to contrasting edaphic conditions identified using landscape scale metagenomics

Ashish A. Malik[1], Bruce C. Thomson[1], Andrew S. Whiteley[2], Mark Bailey[1] and Robert I. Griffiths[1]

[1]*Centre for Ecology and Hydrology, Wallingford, United Kingdom*
[2]*The University of Western Australia, Crawley, Australia*

Correspondence: Ashish A. Malik ashmalik@ceh.ac.uk or Robert I. Griffiths rig@ceh.ac.uk

Abstract

Environmental factors relating to soil pH are widely known to be important in structuring soil bacterial communities, yet the relationship between taxonomic community composition and functional diversity remains to be determined. Here, we analyze geographically distributed soils spanning a wide pH gradient and assess the functional gene capacity within those communities using whole genome metagenomics. Low pH soils consistently had fewer taxa (lower alpha and gamma diversity), but only marginal reductions in alpha diversity and equivalent gamma diversity. However, coherent changes in the relative abundances of annotated genes between pH classes were identified; with functional profiles clustering according to pH independent of geography. Differences in gene abundances were found to reflect survival and nutrient acquisition strategies, with organic-rich acidic soils harboring a greater abundance of cation efflux pumps, C and N direct fixation systems and fermentation pathways indicative of anaerobiosis. Conversely, high pH soils possessed more direct transporter-mediated mechanisms for organic C and N substrate acquisition. These findings show that bacterial functional versatility may not be constrained by taxonomy, and we further identify the range of physiological adaptations required to exist in soils of varying nutrient availability and edaphic conditions.

Introduction

Understanding the key drivers and distributions of microbial biodiversity from both taxonomic and functional perspectives is critical to understand element cycling processes under different land management and geo-climatic scenarios. Distributed soil surveys have shown strong effects of soil properties on the taxonomic biodiversity of bacterial communities (1–5), and to a lesser degree for other soil microbes such as the fungi and protozoa (6, 7). Particularly for bacteria, soil pH often appears as a strong single correlate of biodiversity patterns. This is either due to the direct effects of acidity, or soil pH representing a proxy for a variety of other factors across soil environmental gradients. Acidic soils generally harbor reduced phylogenetic diversity and are

usually dominated by acidophilic Acidobacterial lineages. Intermediate pH soils (pH 5-7) are generally composed of larger numbers of taxa, often with a few dominant lineages; whereas at neutral pH (typically intensive agricultural soils) a more even distribution of a multitude of taxa is typically observed (1, 2, 8). A fundamental issue to be resolved is whether these differences in taxonomic diversity reflect any changes in functional genetic potential. Many dominating organisms, particularly in the more oligotrophic habitats are difficult to culture, and so we know little of the functional characteristics of these organisms at either the phenotypic or genomic level. These knowledge gaps ultimately limit the utility of taxonomic methods (e.g rRNA based) for further understanding ecosystem function.

Whole genome shotgun sequencing has been applied to elucidate the functional diversity of communities from a range of ecosystems (8–12), and its application potentially provides novel insights into the genetic diversity of biochemical processes occurring in soils as well as permitting ecological investigations of microbes within a functional trait based context. The established relationships between edaphic conditions and taxonomic biodiversity can now be complemented with a concurrent understanding of altered environmental microbial physiology and metabolism, which can provide better understanding of the cycling of elements in soils. In the first instance, we need to identify the range of functional genes present in soils to better elucidate the genetic determinants of relevant environmental processes (for example determining the dominant biochemical pathways most likely to be contributing to fluxes). Second, we need to identify the ecological mechanisms determining how genetic pathways are altered according to environmental change. Addressing these knowledge gaps leads to an understanding of how bacterial diversity in the environment relates to functional potential, and importantly how environmental change can impair or enhance soil functionality.

Here we seek to test whether known differences in taxonomic diversity and composition are reflected in functional gene profiles, by implementing metagenomic sequencing of geographically dispersed soils at opposing ends of a temperate bacterial diversity gradient. We chose to sequence 4 low and 4 high pH soils which had previously been collected as part of a national survey of Britain (**figure 1**) and were known to comprise low and high taxonomic diversity respectively (7). In addition to assessing richness effects, we seek to explore change in specific functional gene abundances in order to elucidate the physiological constraints acting on different soil systems and identify variance in functional pathways of relevance to soil biogeochemical cycling.

Results and discussion

Differences in taxonomic richness are not reflected in functional richness

Four geographically distributed soils exhibiting similar pH were selected to represent each of the low and high pH classes based on our previous work. These encompassed a variety of different habitats (**table 1**), with the low pH soils typically being found at higher altitude and possessing higher organic matter and moisture content consistent with broader patterns across Britain. Amplicon sequencing confirmed that the microbial taxonomic richness differed

between the two soil pH groups, with both alpha and gamma diversity being higher in the high pH soil communities (**figure 2a**). We then sought to examine whether the richness of annotated genes from metagenomic sequencing was also less in low pH soils. Metagenomic sequencing using the Roche 454 platform resulted in a total of 4.9 million quality filtered reads across all samples; with 0.4 to 0.7 million reads per sample (**table 1**). An average of 8.3±0.4 % of sequences failed to pass the MG-RAST QC pipeline. Following gene annotation using MG-RAST using default stringency criteria, the percentage of reads annotated to predicted proteins ranged from 63.3 to 73.7 % across samples (average=68.9 %). The majority of annotations were assigned to bacterial taxa (94.7±1.6% in low pH soils, 96.6±0.3% in high pH soils); with only a small proportion being eukaryotic (3±1.2% in low pH soils, 1.8±0.1% in high pH soils) and archaeal (2.2±1.4% in low pH soils, 1.4±0.3% in high pH soils). The marked low proportion of fungi contributing to the soil DNA pool was also reflected in the RDP annotation of ribosomal RNA reads (not shown). The metagenomic guanine plus cytosine content (GC%) was lower in acidic anaerobic soils (**table 1**), corroborating previous evidence suggesting a link between aerobiosis and GC% in prokaryotes (13).

Mean richness (alpha diversity) of functional genes was 4153 and 3915 in high and low pH soils, respectively (**figure 2b**); though this difference was not statistically significant (t-test, $p=0.12$). This was in contrast to the amplicon data, where richness was consistently higher across all samples at high pH (t-test, $p<0.05$). No correlative associations were observed between functional and taxonomic richness, in contrast to one previous study across a range of prairie grasslands (8). Importantly however, we found that whilst the accumulation of taxon richness over sites accentuated the differences in taxonomic diversity, this was not true for functional richness where 4 low pH soils possessed equivalent total functional diversity to 4 high pH soils. Thus, it is clear that whilst taxonomic diversity may be restricted by low pH related factors, functional diversity across multiple samples can be maintained through higher between-site variance (beta diversity) possibly mediated through enhanced metabolic versatility within acidophilic taxa.

Large differences in relative gene abundance between low and high pH soils

Since a lack of difference in functional richness does not mean that soils are functionally similar across the pH gradient, we next sought to assess differences in relative abundances of functional genes. **Figure 3** shows the overall abundance of genes classified at the broadest level (level 1 subsystems classification). Despite similarities in the abundance ranked order of hierarchically classified genes, a number of notable differences between soils of different pH were immediately apparent for several relevant functional categories, such as amino acid cycling, respiration, membrane transport, stress and virulence-disease-defense. However, a lack of difference at this level (notable for carbohydrate and nitrogen metabolism) could be due to significant differences within higher level functional categories. To address this, we next performed a multivariate assessment of gene composition classified at the level of function. This revealed large differences between soils of different pH; indicating that the pH defined communities shared more similar functional gene profiles independent

of their geographical origin (**figure 4a**, anosim $p<0.05$). The analyses based on relative abundance therefore shows that functional genes were differentially abundant in opposing soil pH classes; and in combination with higher functional beta diversity at low pH explains the equivalence of alpha diversity metrics.

Indicator gene analysis (Indval) was then used to define and explore the characteristic genes contributing to the differences between the pH defined soils. Of a total of 6194 annotated genes, 206 and 322 significant functional indicators were found for the low and high pH soils, respectively. A network depicting only strong positive correlations (>0.9) between genes across all samples revealed an expected lack of connectedness between opposing indicators, and in general low pH indicator genes were less correlated than high pH indicators reflecting the greater magnitude of functional beta diversity across low pH soils (**figure 4b**). To examine the functional identity of indicators, we constructed a circular plot (**figure 5**) at the functional gene level of classification, omitting rarer genes for ease of presentation (<50 reads across all samples), and labelling the indicators using the respective level 3 classification. A full table of all gene abundances, their classification and indicator designation are provided in the supplementary material (file S1). The following subsections discuss some relevant indicators of low and high pH soils. Whilst this discussion is by no means exhaustive, we divide the discussion into two sections based on physiological processes for survival and nutrient capture; and metabolism.

Variable physiological strategies for survival and nutrient acquisition

The indicator analysis reveals an array of interlinked physiological adaptations to life at opposing ends of the soil environmental spectrum (**figure 5**). Firstly, with respect to cellular physiology, the high pH soils possessed a greater abundance of ABC transporters relatable to nutrient acquisition (14). In addition to the abundant amino acid, peptide and tricarboxylate transporter indicators (level 1 classification: membrane transport), numerous other transporters were significantly enriched at high pH though under different subsystems classifications. These included the majority of carbohydrate indicators for mono and disaccharide uptake, as well as other transporters for inorganic sulphate, cofactors, polyamines, ammonia/nitrate, potassium uptake proteins, and high affinity phosphate transporters (15). Interestingly, transporters for iron acquisition as well as osmoprotection were also evident at high pH, possibly reflecting low moisture and iron availability in the selected high pH soils. Together these findings indicate that the high pH soils can be distinguished functionally from low pH soils on the basis of a greater abundance of transporters for the direct uptake of available substrates and cofactors required for growth.

The membrane transport related indicators of low pH soils included only two low abundance genes related to nutrient acquisition (phosphate and glucose, not indicated in figure 5 due to low abundance), but a number of genes linked to metal acquisition (ybh, MntH, HoxN/HupN/NixA) and protein secretion (siderophores and extracellular enzymes).

Relatedly a number of membrane proteins (mainly proton antiporters) for the efflux of antibiotics and toxic compounds were characteristic of low pH soils (ACR3, BlaR1/MecR1, CusA, CzcC, MerR, MacB, NodT, MdtB, MdtC), but annotated under "Virulence, Disease and Defense". Indeed, the gene for cation efflux protein cusA related to cobalt-cadmium-zinc resistance was one of the more abundant genes across all metagenomes but was significantly enriched in low pH soils. Given the nature and location of the low pH soils examined it is unlikely these represent adaptations to either severe metal contamination or anthropogenic sources of antibiotics, but rather are related to pH enhanced reactivity of toxicants (e.g. metal ion solubility generally increases with decreasing pH) and the alleviation of acid stress through membrane efflux (16–20). It is also conceivable that they may be required for metal import for a variety of metal necessitating enzymes (see below) since many are proton/cation antiporters. Supporting this, another strong indicator of low pH was a potassium transporting ATPase gene (level 3 class: potassium homeostasis) coding for a membrane protein responsible for exchange of H^+ and K^+ ions across the plasma membrane, suggesting the coupling of acid stress response with elemental acquisition. As an aside, we were unable to find any studies that have examined acid soils for their potential as a source of antibiotic resistance, but our findings implicate adaptation to acidity or anaerobiosis as a possible factor underlying natural resistance to antibiotics (21, 22). Other notable indicators of low pH included chemotaxis and motility genes, plausible given the higher moisture contents of these soils (18). In total, these results identify that in the acidic soils considerable energy investments must be made in cellular processes to survive in an acid stressed, oxygen limited and low nutrient environment.

Metabolic potential of contrasting pH soil communities

Carbohydrate processing was one of the most abundant broad classes of annotated functional processes and within this level 2 class serine-glyoxylate cycle, sugar utilization and TCA cycle genes were identified as the most abundant. Despite the expected conservation of many key metabolic functions, a number of notable indicators were found (**figure 5**). Several genes differed for the processing of mono and oligosaccharides, with a number of genes for L-rhamnose and fructose utilization being of greater abundance at high pH; whereas several extracellular enzyme coding genes for glucosidase, beta galactosidase, mannosidase and hexosamidase were elevated at low pH. With respect to the central carbon metabolism, certain components of the Entner-Doudoroff pathway were reduced in abundance in low pH soils (gluconate dehydratase, gluconolactonase); but by far the most abundant and significant low pH indicator was the xylulose 5-phosphate phosphoketolase gene. This gene along with another strong low pH indicator – transketolase – is part of the pentose phosphate pathway, a metabolic pathway parallel to glycolysis yielding pentose sugars and reducing agents. Though not previously considered largely with respect to soil functionality, this gene was generally of high abundance across all soils. However, novel position-specific isotope tracing experiments have recently provided functional evidence of the significance of the pentose phosphate pathway in soil C cycling (23). Its significance here in low pH organic soils

remains to be questioned, although we note it is involved in fermentative processes and therefore would be expected to occur more frequently in these more anaerobic soils (24). Further evidence of an increase in fermentative processes in low pH soils was seen with a number of indicator genes for pyruvate:ferredoxin oxidoreductase, lactate fermentation (carbohydrate metabolism) and a wide range of hydrogenases (25, 26). Several recent studies have demonstrated that these hydrogenases are widespread in soil microorganisms (25, 27, 28), and they may also represent another means of consuming protons (29). Among these were cytoplasmic NAD-reducing hydrogenases that may be linked to respiration and fermentation through their NADH generation capabilities (30, 31). The nickel transporters (HoxN/HupN/NixA family), identified earlier may also be linked to this process since nickel is required for the metal center of NiFe hydrogenases (32, 33). Together, the metagenomic data suggest that low pH soil communities harbor adaptive physiological strategies of using molecular hydrogen oxidation and coupling it with respiration and fermentation to generate energy (34).

Whilst representing only a small proportion of these metagenomes, a number of nitrogen metabolism associated genes differed significantly between the soils of different pH. Nitrogen fixation genes were consistently strong indicators of low pH soils, with a number of genes found for molybdenum dependent nitrogenases (**figure 5**). Indeed, some of the nitrogenase indicators were unique in being universally present at low pH but entirely absent in the high pH soils. These findings indicate that microbial N input into soils either through symbiotic or non-symbiotic routes may be relatively larger in acidic soils, and there was evidence from the taxonomic assignments that the Bradyrhizobia may play a key role in this process (not shown). Low pH soils are characterized by decreased decomposition rates that could reduce the available N in soils necessitating microbial N fixation (35–37). The coupling of N fixation with abundant hydrogenases may also represents an efficient system for recycling the H_2 produced in N_2 fixation, minimizing the loss of energy (32, 38). High pH soils showed significantly more genes linked to nitrate/nitrite ammonification, ammonia assimilation, and denitrification (**figure 5**). It was interesting to note that no nitrification genes were detected given the level of coverage of these metagenomes and the inaccuracy in read assignment to subsystems. The dominant pathways appeared to be related to ammonification (nitrate reduction to ammonia) and ammonia assimilation, and these were more abundant in high pH soils (39). Relatedly, a number of notable indicators of high pH soils were for the degradation of amino acids and derivatives, such as arginine, ornithine, polyamines, urea and creatine. Coupled with greater abundance of amino acid and peptide transporters, these observations infer an increased reliance on scavenging N enriched compounds originating from biotic inputs in high pH soils (20, 40, 41). Conversely, the lower abundance of these genes at low pH may indicate reduced bioavailability of these compounds possibly due to increased soil adsorption with greater cation exchange capacity (42).

Conclusions

Our study shows that despite large differences in the taxonomic diversity of bacteria known to exist across soil environmental gradients, there was little

evidence to suggest this results in large differences in the diversity of functional genes. Rather, low and high pH soils differed in the relative abundance of specific functional genes, and these indicator genes reflected differences in survival and nutrient acquisition strategies caused through adaptation to different environments. For the low pH soils, there were a number of abundant functional genes that highlight the importance of varied biochemical and physiological processes which we infer to be important adaptations to life in acidic, wet and oxygen limited environments. Indeed our results highlight the coupled action of acidity and anaerobiosis in mediating bacterial functional responses (22). In such soils a considerable investment of energy must be made into complex processes for capturing nutrients and energy, as well as stress responses caused by both the exterior environment and cellular metabolism. In combination, this is reflected by a greater abundance of cation efflux pumps, C and N acquiring systems such as direct fixation, and fermentation (**figure 6**). Higher pH soils conversely possessed more direct transporter mediated mechanisms for C substrate acquisition together with numerous indicators of organic N acquisition and consequent cycling (**figure 6**).

In identifying specific indicators across the gradient, we acknowledge some limitations to the metagenomic approach. Firstly, the analyses are reliant on accurate functional assignment of reads, and there are recognized issues with respect to both bioinformatic annotation and the experimental assignment of a sequence to a single function (43). For this reason, we have focused our discussion on the more prevalent indicators represented by a variety of individual functional gene categories. An additional concern is that observed changes in relative gene abundance could be simply due to change in an unrelated gene (44). Whilst various methods for standardizing reads have been applied (such as relating abundances to rRNA genes, or calculating proportions within discrete subsystems) these approaches are not entirely without scrutiny. We prefer to consider the analyses akin to a mass balance – whereby the relative abundances of genes reflect the proportion of investment made for a given amount of nutrient into different proteins. The relevance for function at the ecosystem scale is a separate line of enquiry necessitating process measurement and assessment of biomass size, and we envisage metagenomic studies will provide more relevant functional targets.

In conclusion, we identify that considerable metabolic diversity and variability can exist within communities of environmentally constrained taxonomic diversity. Our intent at focusing analyses at broad ends of the pH spectrum was inspired to encompass an assessment at the extremes of soil functionality. In doing so, we make available sequence datasets which may be useful to others looking to assess the diversity of specific functional genes for a wide range of soil processes across a range of soils. Furthermore, we envisage that the indicators identified here, whilst being from an extreme soil contrast, may also be relevant at local scales for understanding more subtle alterations in soil function. For example, more "natural" soils in temperate climates typically store more carbon, tend towards acidity, and have increased moisture retention; whereas human agriculture forces soils to neutrality and depletes soil carbon and moisture. Our results identifying differential nutrient acquisition and

alterations in metabolic mechanisms are likely to be widely relevant in understanding the balance of energy and C storage mechanisms under altered land management; as well as permitting future design of smarter systems for efficient soil nutrient capture and recycling.

Materials and Methods

Characteristics of soil examined and location of sampling sites are shown in **Table 1** and **Fig. 1**, and full details of the sampling, nucleic acid extraction and taxonomic analyses are provided in our previous manuscripts (2, 7). Eight soils were selected for detailed metagenomic analysis on the basis of soil pH alone and are representative of the extremes of both the soil environmental conditions and bacterial diversity range encountered across Britain. Bacterial communities were characterized using tagged amplicon sequencing as previously described (7) using primers 28F / 519R, and sequencing on the 454 pyrosequencing platform through a commercial provider. Raw sequences from amplicon sequencing were analyzed using QIIME, using UCLUST to generate OTU's at 97% sequence similarity. For whole genome metagenomic sequencing, three μg of total DNA from each sample was submitted to the NERC Biomolecular Analyses Facility (Liverpool, UK) with two samples per run on the 454 platform (sequencing statistics in **Table 1**). Resulting sequences from metagenomic analysis were annotated with the Metagenomics Rapid Annotation using Subsystems Technology (MG-RAST) server version 4.0 (45). Functional classification was performed using the SEED Subsystems database with a maximum e-value cut-off of 10^{-5}, minimum identity cut-off of 60% and minimum length of sequence alignment of 15 nucleotides. Gene abundance tables derived from MG-RAST were imported into R for downstream analyses. Rarefaction, species diversity accumulation, ordinations, and statistical analyses were performed using the vegan package (46) under the R environment software 2.14.0 (R Development Core Team 2011).

Funding information

This project was funded by the UK Natural Environment Research Council (standard grant NE/E006353/1 to RIG, ASW & MB; and Soil Security grant NE/M017125/1 to RIG). AAM has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant No 655240.

References

1. **Fierer N**, **Jackson RB**. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103**:626–631.

2. **Griffiths RI**, **Thomson BC**, **James P**, **Bell T**, **Bailey M**, **Whiteley AS**. 2011. The bacterial biogeography of British soils. Environ Microbiol **13**:1642–1654.

3. **Tripathi BM**, **Kim M**, **Singh D**, **Lee-Cruz L**, **Lai-Hoe A**, **Ainuddin AN**, **Go R**, **Rahim RA**, **Husni MHA**, **Chun J**, **Adams JM**. 2012. Tropical Soil Bacterial Communities in Malaysia: PH Dominates in the Equatorial Tropics Too. Microb Ecol **64**:474–484.

4. **Liu J**, **Sui Y**, **Yu Z**, **Shi Y**, **Chu H**, **Jin J**, **Liu X**, **Wang G**. 2014. High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. Soil Biol Biochem **70**:113–122.

5. **Lauber CL**, **Hamady M**, **Knight R**, **Fierer N**. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol **75**:5111–

5120.

6. **Tedersoo L**, **Bahram M**, **Põlme S**, **Kõljalg U**, **Yorou NS**, **Wijesundera R**, **Ruiz LV**, **Vasco-Palacios AM**, **Thu PQ**, **Suija A**, **Smith ME**, **Sharp C**, **Saluveer E**, **Saitta A**, **Rosas M**, **Riit T**, **Ratkowsky D**, **Pritsch K**, **Põldmaa K**, **Piepenbring M**, **Phosri C**, **Peterson M**, **Parts K**, **Pärtel K**, **Otsing E**, **Nouhra E**, **Njouonkou AL**, **Nilsson RH**, **Morgado LN**, **Mayor J**, **May TW**, **Majuakim L**, **Lodge DJ**, **Lee SS**, **Larsson K-H**, **Kohout P**, **Hosaka K**, **Hiiesalu I**, **Henkel TW**, **Harend H**, **Guo L**, **Greslebin A**, **Grelet G**, **Geml J**, **Gates G**, **Dunstan W**, **Dunk C**, **Drenkhan R**, **Dearnaley J**, **De Kesel A**, **Dang T**, **Chen X**, **Buegger F**, **Brearley FQ**, **Bonito G**, **Anslan S**, **Abell S**, **Abarenkov K**. 2014. Global diversity and geography of soil fungi. Science (80-) **346**.

7. **Dupont A**, **Griffiths RI**, **Bell T**, **Bass D**. 2016. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. Environ Microbiol **18**:2010–24.

8. **Fierer N**, **Ladau J**, **Clemente JC**, **Leff JW**, **Owens SM**, **Pollard KS**, **Knight R**, **Gilbert J a**, **McCulley RL**. 2013. Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science **342**:621–4.

9. **Fierer N**, **Leff JW**, **Adams BJ**, **Nielsen UN**, **Bates ST**, **Lauber CL**, **Owens SM**, **Gilbert J a.**, **Wall DH**, **Caporaso JG**. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A **109**:21390–5.

10. **Hultman J**, **Waldrop MP**, **Mackelprang R**, **David MM**, **McFarland J**, **Blazewicz SJ**, **Harden J**, **Turetsky MR**, **McGuire AD**, **Shah MB**, **VerBerkmoes NC**, **Lee LH**, **Mavrommatis K**, **Jansson JK**. 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. Nature **521**:208–212.

11. **Myrold DD**, **Zeglin LH**, **Jansson JK**. 2014. The Potential of Metagenomic Approaches for Understanding Soil Microbial Processes. Soil Sci Soc Am J **78**:3–10.

12. **Edwards A**, **Pachebat JA**, **Swain M**, **Hegarty M**, **Hodson AJ**, **L Irvine-Fynn TD**, **E Rassner SM**, **Sattler B**. 2013. A metagenomic snapshot of taxonomic and functional diversity in an alpine glacier cryoconite ecosystem. Environ Res Lett **8**:35003–11.

13. **Naya H**, **Romero H**, **Zavala A**, **Alvarez B**, **Musto H**. 2002. Aerobiosis Increases the Genomic Guanine Plus Cytosine Content (GC%) in Prokaryotes. J Mol Evol **55**:260–264.

14. **Davidson AL**, **Dassa E**, **Orelle C**, **Chen J**. 2008. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. Microbiol Mol Biol Rev **72**:317–364.

15. **Jiménez DJ**, **Chaves-Moreno D**, **van Elsas JD**. 2015. Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. Sci Rep **5**:13845.

16. **Outten FW**, **Huffman DL**, **Hale JA**, **O'Halloran T V.** 2001. The Independent cue and cus Systems Confer Copper Tolerance during Aerobic and Anaerobic Growth in Escherichia coli. J Biol Chem **276**:30670–30677.

17. **Jakob K**, **Satorhelyi P**, **Lange C**, **Wendisch VF**, **Silakowski B**, **Scherer S**, **Neuhaus K**. 2007. Gene expression analysis of Corynebacterium glutamicum subjected to long-term lactic acid adaptation. J Bacteriol **189**:5582–5590.

18. **Ryan D**, **Pati NB**, **Ojha UK**, **Padhi C**, **Ray S**, **Jaiswal S**, **Singh GP**, **Mannala GK**, **Schultze T**, **Chakraborty T**, **Suar M**. 2015. Global transcriptome and mutagenic analyses of the acid tolerance response of Salmonella enterica serovar typhimurium. Appl Environ Microbiol **81**:8054–8065.

19. **Lund P**, **Tramonti A**, **De Biase D**. 2014. Coping with low pH: molecular strategies in

neutralophilic bacteria. FEMS Microbiol Rev **38**:1091–1125.

20. **Wilks JC**, **Kitko RD**, **Cleeton SH**, **Lee GE**, **Ugwu CS**, **Jones BD**, **Bondurant SS**, **Slonczewski JL**. 2009. Acid and base stress and transcriptomic responses in Bacillus subtilis. Appl Environ Microbiol **75**:981–990.

21. **Alvarez-Ortega C**, **Olivares J**, **Martínez JL**. 2013. RND multidrug efflux pumps: What are they good for? Front Microbiol **4**:1–11.

22. **Hayes ET**, **Wilks JC**, **Sanfilippo P**, **Yohannes E**, **Tate DP**, **Jones BD**, **Radmacher MD**, **BonDurant SS**, **Slonczewski JL**. 2006. Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in Escherichia coli K-12. BMC Microbiol **6**:89.

23. **Apostel C**, **Dippold M**, **Kuzyakov Y**. 2015. Biochemistry of hexose and pentose transformations in soil analyzed by position-specific labeling and 13C-PLFA. Soil Biol Biochem **80**:199–208.

24. **Sanchez B**, **Zuniga M**, **Gonzalez-Candelas F**, **de los Reyes-Gavilan CG**, **Margolles A**. 2010. Bacterial and eukaryotic phosphoketolases: phylogeny, distribution and evolution. J Mol Microbiol Biotechnol **18**:37–51.

25. **Greening C**, **Biswas A**, **Carere CR**, **Jackson CJ**, **Taylor MC**, **Stott MB**, **Cook GM**, **Morales SE**. 2015. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME J **10**:1–17.

26. **Berney M**, **Greening C**, **Conrad R**, **Jacobs WR**, **Cook GM**. 2014. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. Proc Natl Acad Sci U S A **111**:11479–84.

27. **Constant P**, **Chowdhury SP**, **Hesse L**, **Pratscher J**, **Conrad R**. 2011. Genome data mining and soil survey for the novel group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity H 2-oxidizing bacteria. Appl Environ Microbiol **77**:6027–6035.

28. **Baba R**, **Kimura M**, **Asakawa S**, **Watanabe T**. 2014. Analysis of [FeFe]-hydrogenase genes for the elucidation of a hydrogen-producing bacterial community in paddy field soil. FEMS Microbiol Lett **350**:249–256.

29. **Noguchi K**, **Riggins DP**, **Eldahan KC**, **Kitko RD**, **Slonczewski JL**. 2010. Hydrogenase-3 contributes to anaerobic acid resistance of escherichia coli. PLoS One **5**:27–29.

30. **Lauterbach L**, **Lenz O**, **Vincent KA**. 2013. H2-driven cofactor regeneration with NAD(P)+-reducing hydrogenases. FEBS J **280**:3058–3068.

31. **Burgdorf T**, **Linden E Van Der**, **Bernhard M**, **Yin QY**, **Back JW**, **Hartog AF**, **Anton O**, **Koster CG De**, **Albracht SPJ**, **Friedrich B**, **Muijsers AO**. 2005. The Soluble NAD + -Reducing [NiFe] -Hydrogenase from Ralstonia eutropha H16 Consists of Six Subunits and Can Be Specifically Activated by NADPH The Soluble NAD ƨ -Reducing [NiFe] -Hydrogenase from Ralstonia eutropha H16 Consists of Six Subunits and Ca. J Bacteriol **187**:3122–3132.

32. **Bothe H**, **Schmitz O**, **Yates MG**, **Newton WE**. 2010. Nitrogen fixation and hydrogen metabolism in cyanobacteria. Microbiol Mol Biol Rev **74**:529–551.

33. **Eitinger T**, **Mandrand-Berthelot M-A**. 2000. Nickel transport systems in microorganisms. Arch Microbiol **173**:1–9.

34. **Greening C**, **Carere CR**, **Rushton-Green R**, **Harold LK**, **Hards K**, **Taylor MC**, **Morales SE**, **Stott MB**, **Cook GM**. 2015. Persistence of the dominant soil phylum Acidobacteria by trace gas scavenging. Proc Natl Acad Sci U S A **112**:10497–502.

35. **Hsu SF**, **Buckley DH**. 2009. Evidence for the functional significance of diazotroph community structure in soil. ISME J **3**:124–136.

36. **Silver WL**, **Thompson AW**, **Reich A**, **Ewel JJ**, **Firestone MK**. 2005. Nitrogen cycling in tropical plantation forests: Potential controls on nitrogen retention. Ecol Appl **15**:1604–1614.

37. **Cobo-Diaz JF**, **Fernandez-Gonzalez AJ**, **Villadas PJ**, **Robles AB**, **Toro N**, **Fernandez-Lopez M**. 2015. Metagenomic Assessment of the Potential Microbial Nitrogen Pathways in the Rhizosphere of a Mediterranean Forest After a Wildfire. Microb Ecol **69**:895–904.

38. **Ng G**, **Tom CGS**, **Park AS**, **Zenard L**, **Ludwig RA**. 2009. A novel endo-hydrogenase activity recycles hydrogen produced by Nitrogen fixation. PLoS One **4**:e4695.

39. **Giles M**, **Morley N**, **Baggs EM**, **Daniell TJ**. 2012. Soil nitrate reducing processes - Drivers, mechanisms for spatial variation, and significance for nitrous oxide production. Front Microbiol **3**:1–16.

40. **Wilkinson A**, **Hill PW**, **Farrar JF**, **Jones DL**, **Bardgett RD**. 2014. Rapid microbial uptake and mineralization of amino acids and peptides along a grassland productivity gradient. Soil Biol Biochem **72**:75–83.

41. **Apostel C**, **Dippold M**, **Glaser B**, **Kuzyakov Y**. 2013. Biochemical pathways of amino acids in soil: Assessment by position-specific labeling and 13C-PLFA analysis. Soil Biol Biochem **67**:31–40.

42. **Cleve K Van**, **Dyrness CT**, **Marion GM**, **Erickson R**. 1993. Control of soil development on the Tanana River floodplain, interior Alaska. Can J For Res **23**:941–955.

43. **Huberts DHEW**, **van der Klei IJ**. 2010. Moonlighting proteins: An intriguing mode of multitasking. Biochim Biophys Acta - Mol Cell Res **1803**:520–525.

44. **Lovell D**, **Pawlowsky-Glahn V**, **Egozcue JJ**, **Marguerat S**, **Bähler J**. 2015. Proportionality: A Valid Alternative to Correlation for Relative Data. PLoS Comput Biol **11**:1–12.

45. **Meyer F**, **Paarmann D**, **D'Souza M**, **Olson R**, **Glass EM**, **Kubal M**, **Paczian T**, **Rodriguez A**, **Stevens R**, **Wilke A**. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9**:386.

46. **Oksanen J**, **Blanchet FG**, **Kindt R**, **Legendre P**, **Minchin PR**, **O'Hara RB**, **Simpson GL**, **Solymos P**, **Stevens MHH**, **Wagner H**. 2015. vegan: Community Ecology Package. R package version 2.3-0.

Table 1: Summary of soil and metagenomic characteristics.

Sample	pH	Land use type	Organic matter (%)	Moisture (%)	No. reads	G:C content
4475877	4.5	Heath	21.2	56	400965	60
4475885	4.3	Acid grassland	70.4	82	612521	62
4475892	4.1	Acid grassland	47.8	72	647598	61
4475897	4.4	Coniferous woodland	88.8	90	645976	59
4475881	8.3	Hedgerow	24.1	48	556279	63
4475888	8.4	Improved grassland	6.4	25	621879	63
4475891	8	Arable	5.1	21	679130	64
4475893	8.5	Improved grassland	8.9	33	722025	63

Figure legends

Figure 1: Geographically distributed soils from a range of habitats sampled at opposing ends of a landscape pH gradient. The sampling locations are displayed over a soil pH map of Britain derived from the UK Soils portal (ukso.org)

Figure 2: Within site and across site taxonomic and functional richness represented by site based accumulation curves. Standard deviations are calculated from random permutations of the data, with red lines representing low pH and blue lines high pH. (a) Taxonomic richness determined by 16S rRNA sequencing is higher at high pH, both within individual sites (alpha diversity) and accumulated across sites (gamma diversity); (b) richness of annotated functional genes following whole genome sequencing is only marginally lower in low pH at individual sites, and converges across all sites.

Figure 3: Abundances of annotated functional genes classified at the broadest level (level 1 subsystems classification), with total reads standardised across samples to 92442 reads.

Figure 4: (a) Ordination of functional genes (classified at the level of function) using two dimensional NMDS reveals strong clustering of sites by pH irrespective of geographical sampling origin. (b) Network depicting strong (>0.9) positive correlations between annotated functional genes. For clarity, rare genes with less than 10 instances across all samples were omitted. Significant indicators are coloured according to pH class following indicator (indval) analyses.

Figure 5: (a) Bar plot showing the frequency of indicator genes at the broad level 1 classification. (b) Circular plot displaying the identity and abundances of indicator genes for low and high pH soils. Nodes represent individual functional indicators, though are labelled with the more descriptive subsystems level 3 classification, i.e. repeated node labels indicate

different functional indicators within the same level 3 subsystems classification. Node labels are coloured red and blue for particular genes that are significantly more abundant in low or high pH soils, respectively. Line plots represent total abundances of the indicators within the rarefied datasets, and are filled according to pH (red=low, blue=high). The tree depicts the hierarchical subsystems classification, with level 1 classifications being labelled on the internal nodes.

Figure 6: Schematic summarising some of the main physiological differences for survival, nutrient acquisitions and substrate metabolism across the pH gradient, as identified from the indicator analyses. We note that inclusion of a gene in either schematic is based on differences in abundances and does not implicate the presence/absence of a particular pathway across the gradient (refer to figure 5b).

Fig.1

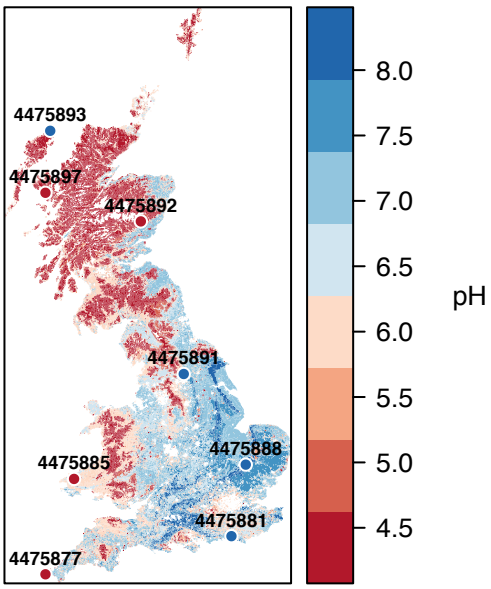

4475893
4475897
4475892
4475891
4475888
4475885
4475881
4475877
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
pH

Fig.2

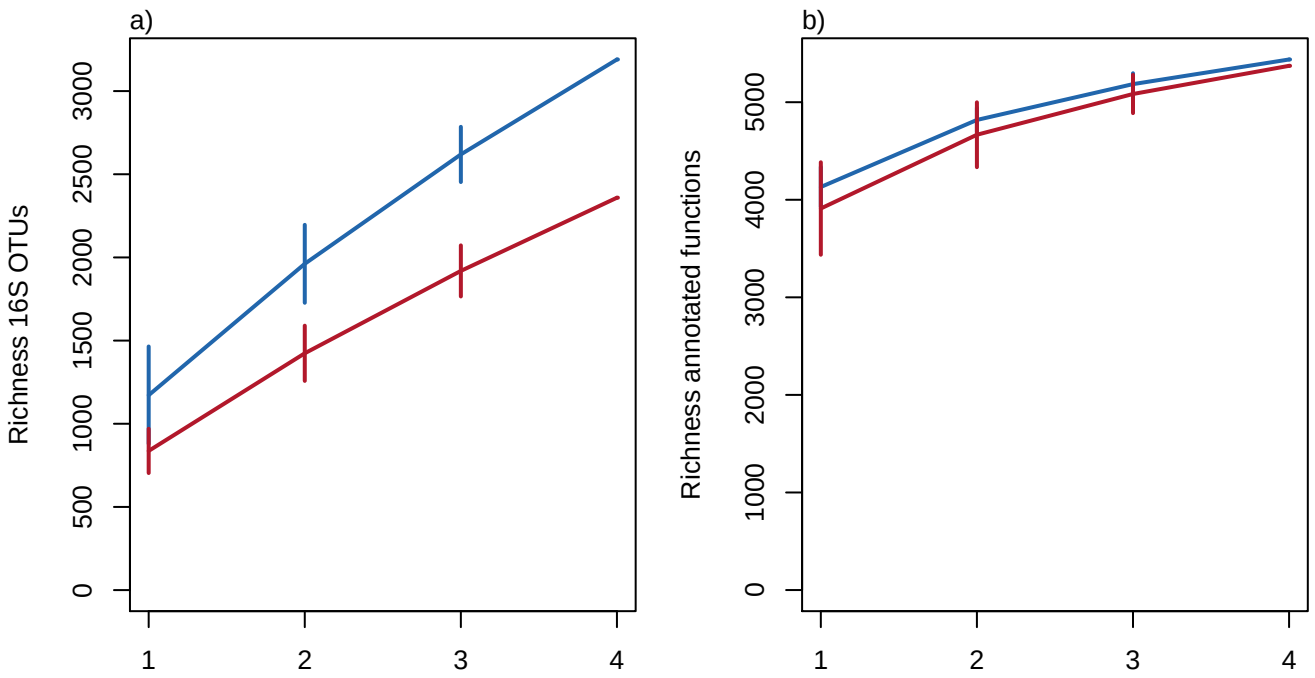

a)
Richness 16S OTUs
0
500
1000
1500
2000
2500
3000
1
2
3
4
b)
Richness annotated functions
0
1000
2000
3000
4000
5000
1
2
3
4

Fig.3

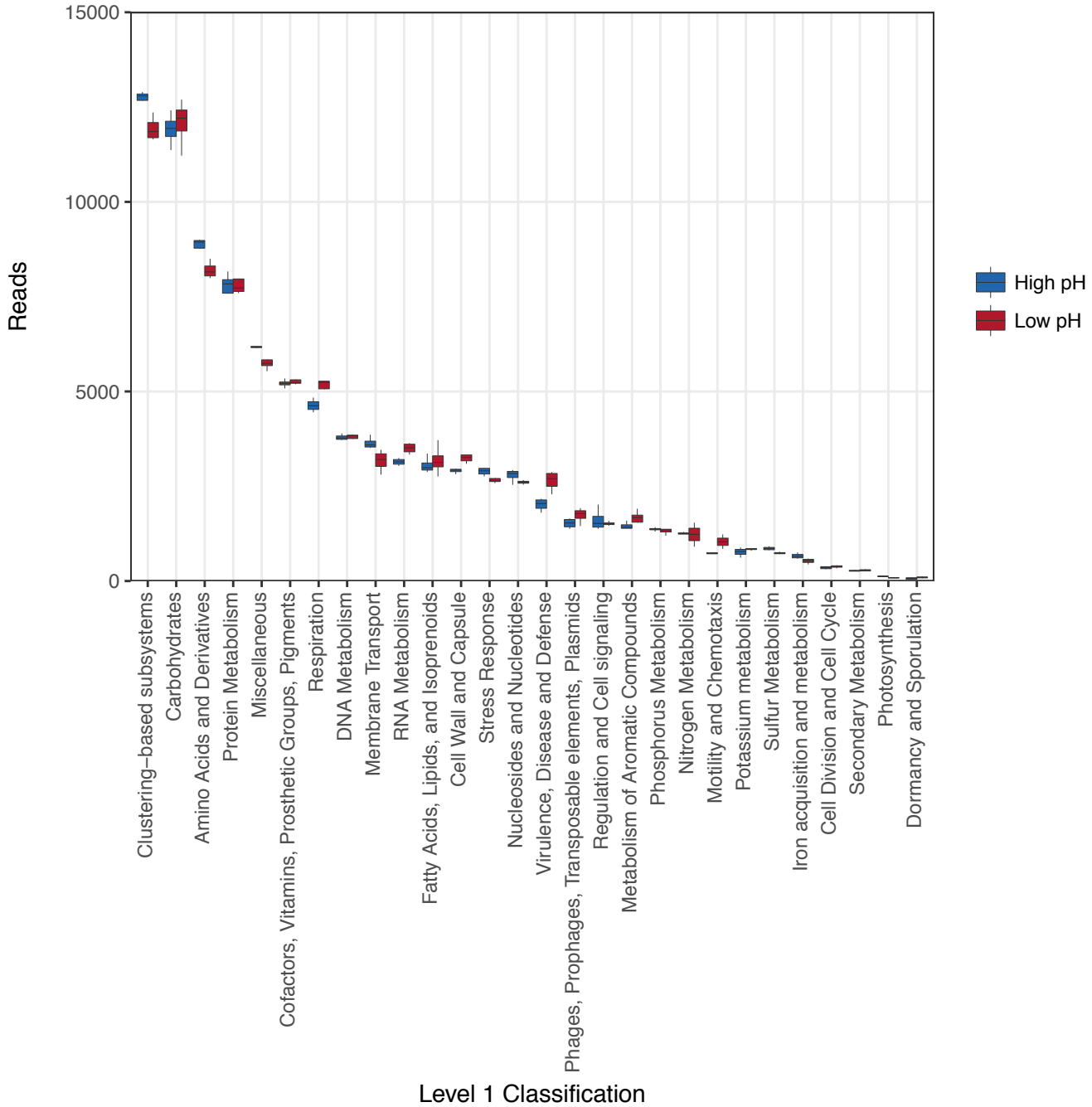

15000
10000
5000
0
Reads
High pH
Low pH
Clustering−based subsystems
Carbohydrates
Amino Acids and Derivatives
Protein Metabolism
Miscellaneous
Cofactors, Vitamins, Prosthetic Groups, Pigments
Respiration
DNA Metabolism
Membrane Transport
RNA Metabolism
Fatty Acids, Lipids, and Isoprenoids
Cell Wall and Capsule
Stress Response
Nucleosides and Nucleotides
Virulence, Disease and Defense
Phages, Prophages, Transposable elements, Plasmids
Regulation and Cell signaling
Metabolism of Aromatic Compounds
Phosphorus Metabolism
Nitrogen Metabolism
Motility and Chemotaxis
Potassium metabolism
Sulfur Metabolism
Iron acquisition and metabolism
Cell Division and Cell Cycle
Secondary Metabolism
Photosynthesis
Dormancy and Sporulation
Level 1 Classification

Fig.4

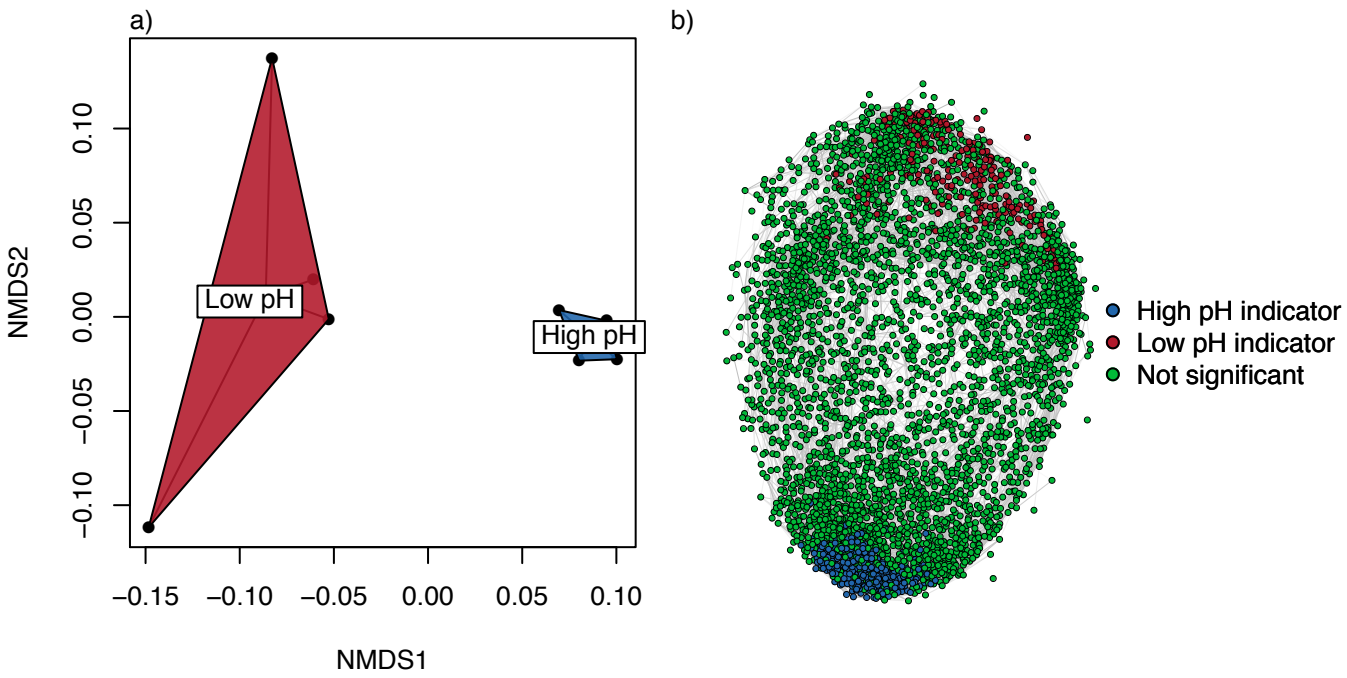
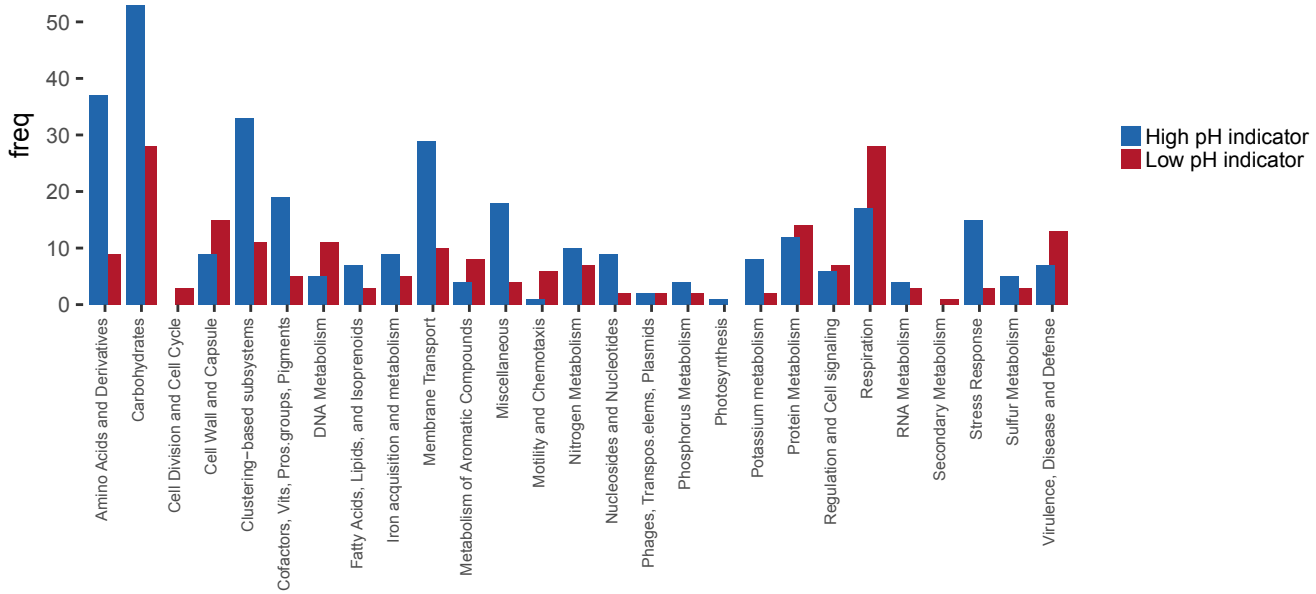

a)
b)
Low pH
High pH
NMDS1
NMDS2
−0.15
−0.10
−0.05
0.00
0.05
0.10
High pH indicator
Low pH indicator
Not significant

Fig.5


freq
50
40
30
20
10
0
High pH indicator
Low pH indicator
Amino Acids and Derivatives
Carbohydrates
Cell Division and Cell Cycle
Cell Wall and Capsule
Clustering−based subsystems
Cofactors, Vits, Pros.groups, Pigments
DNA Metabolism
Fatty Acids, Lipids, and Isoprenoids
Iron acquisition and metabolism
Membrane Transport
Metabolism of Aromatic Compounds
Miscellaneous
Motility and Chemotaxis
Nitrogen Metabolism
Nucleosides and Nucleotides
Phages, Transpos.elems, Plasmids
Phosphorus Metabolism
Photosynthesis
Potassium metabolism
Protein Metabolism
Regulation and Cell signaling
Respiration
RNA Metabolism
Secondary Metabolism
Stress Response
Sulfur Metabolism
Virulence, Disease and Defense

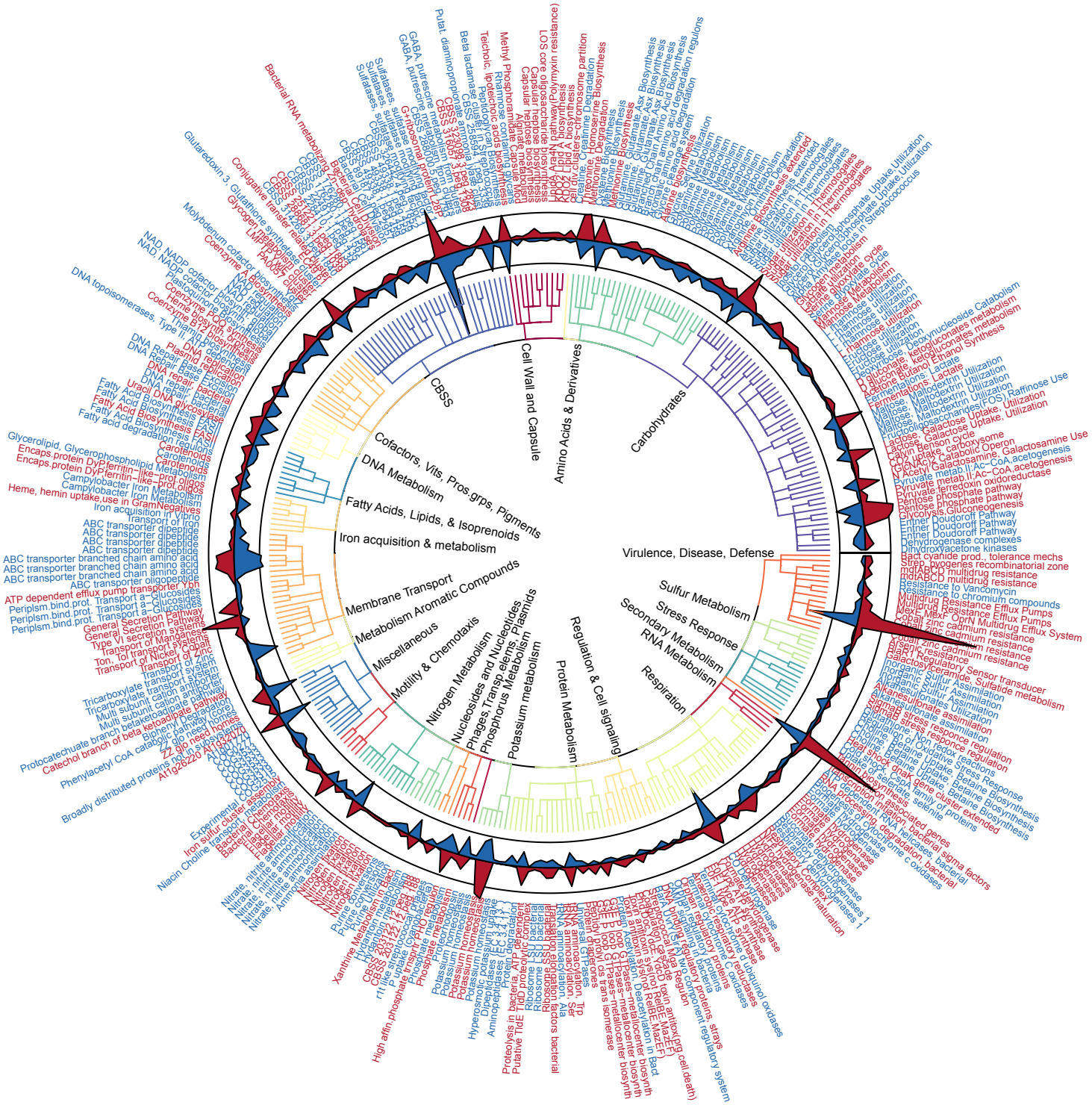
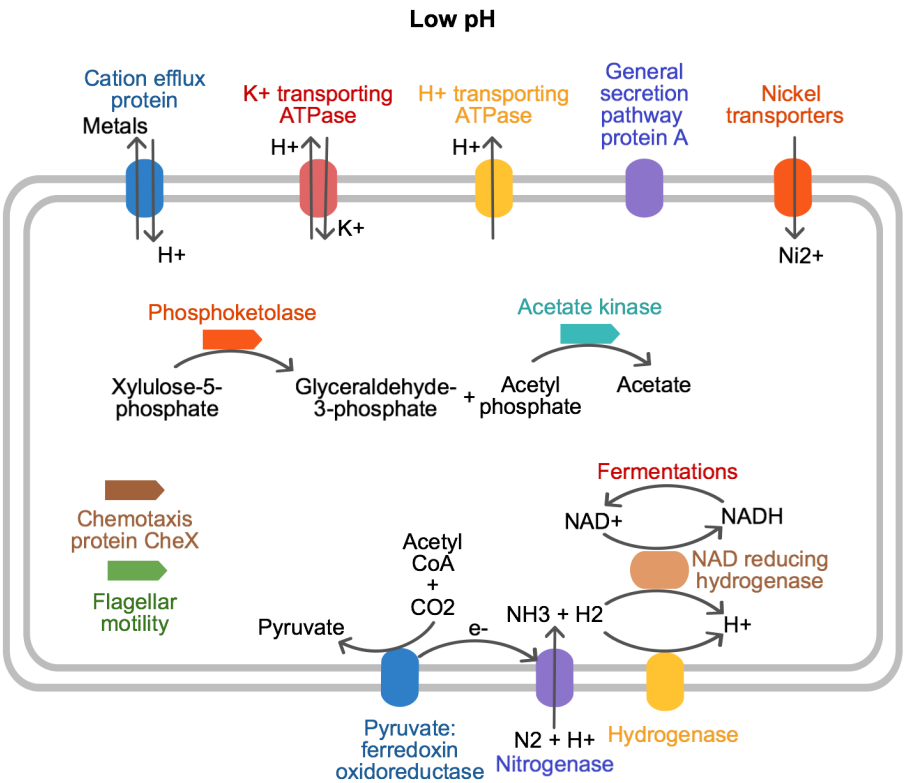
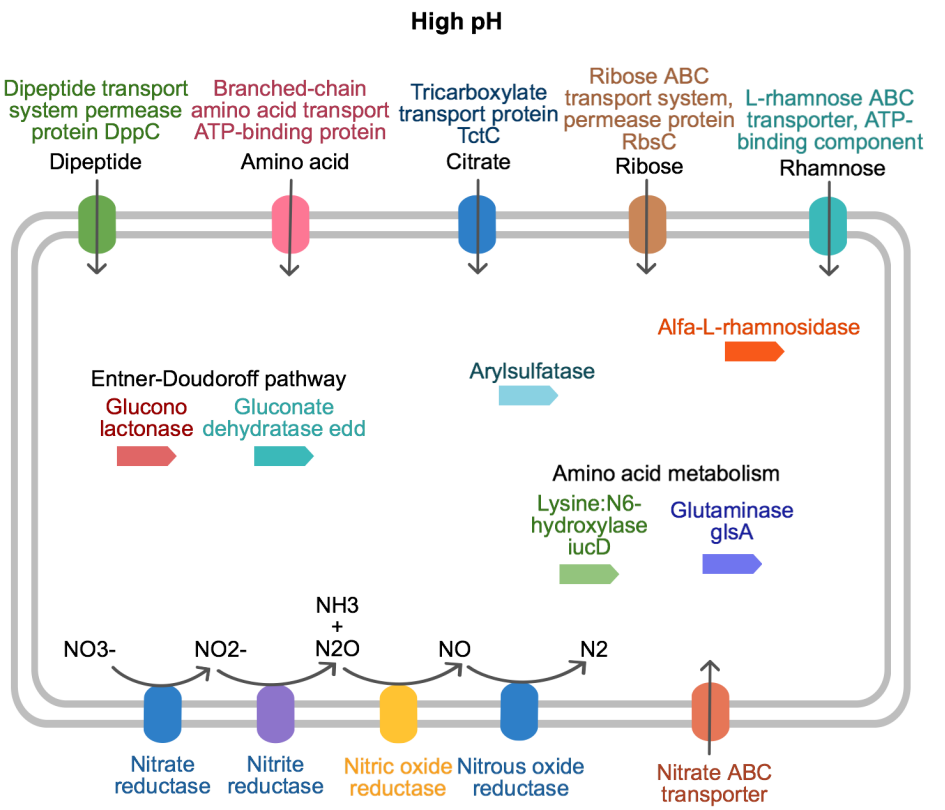

Cell Wall and Capsule
Amino Acids & Derivatives
Carbohydrates
CBSS
Cofactors, Vits, Pros.grps, Pigments
DNA Metabolism
Fatty Acids, Lipids, & Isoprenoids
Iron acquisition & metabolism
Virulence, Disease, Defense
Membrane Transport
Metabolism Aromatic Compounds
Miscellaneous
Motility & Chemotaxis
Nitrogen Metabolism
Nucleosides and Nucleotides
Phages,Transp.elems, Plasmids
Phosphorus Metabolism
Potassium metabolism
Protein Metabolism
Regulation & Cell signaling
Respiration
RNA Metabolism
Secondary Metabolism
Stress Response
Sulfur Metabolism

Fig.6


Low pH
Cation efflux protein
Metals
H+
K+ transporting ATPase
H+
K+
H+ transporting ATPase
H+
General secretion pathway protein A
Nickel transporters
Ni2+
Phosphoketolase
Xylulose-5-phosphate
Glyceraldehyde-3-phosphate
+
Acetate kinase
Acetyl phosphate
Acetate
Fermentations
NAD+
NADH
Chemotaxis protein CheX
Flagellar motility
Acetyl CoA + CO2
NAD reducing hydrogenase
Pyruvate
e-
NH3 + H2
H+
Pyruvate: ferredoxin oxidoreductase
N2 + H+
Nitrogenase
Hydrogenase


High pH
Dipeptide transport system permease protein DppC
Dipeptide
Branched-chain amino acid transport ATP-binding protein
Amino acid
Tricarboxylate transport protein TctC
Citrate
Ribose ABC transport system, permease protein RbsC
Ribose
L-rhamnose ABC transporter, ATP-binding component
Rhamnose
Alfa-L-rhamnosidase
Arylsulfatase
Entner-Doudoroff pathway
Glucono lactonase
Gluconate dehydratase edd
Amino acid metabolism
Lysine:N6-hydroxylase iucD
Glutaminase glsA
NO3-
NO2-
NH3 + N2O
NO
N2
Nitrate reductase
Nitrite reductase
Nitric oxide reductase
Nitrous oxide reductase
Nitrate ABC transporter