The bacterial biogeography of British soils

Robert I. Griffiths,[1]* Bruce C. Thomson,[1]
Phillip James,[1,2] Thomas Bell,[3] Mark Bailey[1] and
Andrew S. Whiteley[1]

[1]Centre for Ecology & Hydrology, Maclean Building,
Benson Lane, Crowmarsh Gifford, Wallingford,
Oxfordshire, OX10 8BB, UK.
[2]School of Biology, Ridley Building, University of
Newcastle upon Tyne, Newcastle upon Tyne, NE1 7RU,
UK.
[3]Department of Zoology, University of Oxford, South
Parks Road, Oxford, OX1 3PS, UK.

Summary

Despite recognition of the importance of soil bacteria to terrestrial ecosystem functioning there is little consensus on the factors regulating belowground biodiversity. Here we present a multi-scale spatial assessment of soil bacterial community profiles across Great Britain (> 1000 soil cores), and show the first landscape scale map of bacterial distributions across a nation. Bacterial diversity and community dissimilarities, assessed using terminal restriction fragment length polymorphism, were most strongly related to soil pH providing a large-scale confirmation of the role of pH in structuring bacterial taxa. However, while α diversity was positively related to pH, the converse was true for β diversity (between sample variance in α diversity). β diversity was found to be greatest in acidic soils, corresponding with greater environmental heterogeneity. Analyses of clone libraries revealed the pH effects were predominantly manifest at the level of broad bacterial taxonomic groups, with acidic soils being dominated by few taxa (notably the group 1 Acidobacteria and Alphaproteobacteria). We also noted significant correlations between bacterial communities and most other measured environmental variables (soil chemistry, aboveground features and climatic variables), together with significant spatial correlations at close distances. In particular, bacterial and plant communities were closely related signifying no strong evidence that soil bacteria are driven by different ecological processes to those governing higher organisms. We conclude that broad scale surveys are useful in identifying distinct soil biomes comprising reproducible communities of dominant taxa. Together these results provide a baseline ecological framework with which to pursue future research on both soil microbial function, and more explicit biome based assessments of the local ecological drivers of bacterial biodiversity.

Introduction

Bacteria constitute a major portion of the biodiversity in soils (Gans *et al.*, 2005; Roesch *et al.*, 2007; Fulthorpe *et al.*, 2008), and play an essential role in maintaining soil processes (Brussaard, 1997; Bardgett *et al.*, 2008), which ultimately affect the functioning of terrestrial ecosystems. An understanding of the factors that influence the biodiversity of soil bacterial communities is needed, first as a framework for determining the roles of different taxa in soils; and also to predict ecosystem responses to a changing environment. From a natural historical perspective, our understanding of how bacterial communities are distributed at landscape scales remains rudimentary (Dequiedt *et al.*, 2009). More broadly, microbial species have been perceived as being ubiquitous and are often assumed to be functionally redundant, leading some researchers to suspect that microbes follow different ecological rules to higher organisms (discussed in Martiny *et al.*, 2006). There is therefore growing wider interest in whether microbial biogeographic patterns differ fundamentally from those of larger organisms (Horner-Devine *et al.*, 2004; Martiny *et al.*, 2006; Prosser *et al.*, 2007).

Of fundamental importance is the basic ecological requirement to determine whether microbes are ubiquitously or randomly distributed (Martiny *et al.*, 2006; Ramette and Tiedje, 2007); and if they are not, to elucidate the relative influence of biotic and abiotic factors in affecting community structure. Previous studies based on pure cultures isolated from soils have typically revealed endemism within defined isolated taxa (Cho and Tiedje, 2000; Ramette and Tiedje, 2007). However, the taxa detected by culturing are known to not reflect the dominant taxa in an environment (Dunbar *et al.*, 1999; Amann and Ludwig, 2000), and molecular methods based on diversity profiling of environmental DNA are more useful if we wish to gain additional information on how the dominant members of bacterial communities differ in composition across landscapes. Using such methods, Fierer and

Jackson (2006) found in 98 soil samples widely distributed across the Americas, that pH was the main driver determining the richness and composition of bacterial taxa, above and beyond other environmental and spatial factors. More recent reports have also documented the overriding effect of pH on soil bacterial communities assessed with molecular community profiling approaches, although they have focussed on discrete geographic regions (Hartman *et al.*, 2008a; Baker *et al.*, 2009; Jesus *et al.*, 2009). To bridge the gap between local and wide-scale assessments, more studies are required encompassing greater spatial sampling across multiple soil biomes, to reveal whether the predictable relationships between total bacterial communities and specific environmental factors are a global feature of bacterial biogeography. Furthermore, some assessment of the taxonomic resolution provided by the 16S based methodologies is required in order to address the conflicting conclusions regarding the stronger spatial effects apparent in the local studies of bacterial isolates.

The relationships observed between soil pH and bacterial community structure have largely been based on correlations of pH and an index of diversity, i.e. a univariate descriptor of the number and proportions of taxa within each sample. Despite some criticism that these measures only reflect the diversity of taxa detectable with a molecular method, and not the entire bacterial community (Bent *et al.*, 2007; Blackwood *et al.*, 2007), simple indices do allow for comparisons of community–environment relationships across studies, and are therefore more suitable than multivariate comparisons of pairwise distances in providing wide-scale synthesis. A unimodal relationship between diversity and pH was observed in the American survey, while a more geographically constrained study of wetland soils did not show the same characteristic drop in diversity at low pH (Hartman *et al.*, 2008a), highlighting that further studies are needed to globally synthesize the precise relationships between bacterial diversity and pH. Additionally, a key issue that has yet to be addressed in studies of soil bacterial diversity is how β diversity, the variability in communities among samples (Whittaker, 1960), changes across dominant soil environmental gradients. Large-scale surveys of β diversity are critical for understanding how bacterial communities respond to environmental factors as communities with similar levels of α diversity might be composed of very different taxa.

Determining general rules governing bacterial biogeography at a global scale therefore requires widespread spatial surveillance of community diversity and composition, coupled with extensive collection of environmental data. We sought to analyse bacterial communities in over 1000 soils collected across Great Britain as part of a nationwide monitoring scheme. The soils were collected from the 2007 Countryside Survey (www.countrysidesurvey.org.uk/); a long-term British monitoring programme incorporating extensive sampling of habitat and soil characteristics at multiple spatial scales. We used a simple molecular community profiling tool [terminal restriction fragment length polymorphism (TRFLP) analysis] and limited sequencing of selected clone libraries (V1–V3 regions of 16S rRNA) to assess the biogeography of soil bacterial communities with respect to three broad and related aims: (i) to address the environmental and spatial determinants of bacterial biogeography at multiple spatial scales across Britain; (ii) To investigate how bacterial diversity and beta diversity vary according to the detected environmental gradients and (iii) to generate a broad-scale map of patterns of bacterial biodiversity at a nationwide scale.

Results

Explaining variance in bacterial community diversity

Simpsons index was used to calculate bacterial α diversity in each of the soil samples ($n = 1010$), and was then used as a response variable in multiple regressions, using each environmental variable as a predictor (for variables see *Experimental procedures*). Soil pH was found to be the strongest predictor of α diversity ($r^2 = 0.41$, $P < 0.0001$, $n = 1010$, Fig. 1), followed by the first axis scores of a plant community ordination ($r^2 = 0.27$, $P < 0.0001$, $n = 1010$). While remaining significant all other variables generally were weakly correlated with r^2 values of less than 0.2. The large effect of plant communities was also reflected when testing the effects of the aggregate vegetation class categorical variables on diversity ($r^2 = 0.29$, $P < 0.0001$, $n = 990$), indicating that typically 'improved' habitats such as agricultural land and grasslands have higher soil pH and consequently higher diversity. As soil pH increased, bacterial diversity between samples became less variable. The larger amount of variation in diversity between low pH soil samples raises the possibility that over larger areas, meta-communities of low pH soils may in fact be more diverse. We tested this by applying additive diversity partitioning (Lande, 1996), to samples grouped into discrete pH classes. While lower pH soils exhibited lower within-sample diversity (α diversity), between-sample variance (β diversity) declined with increasing pH (β diversity v pH: $r^2 = 0.19$). Therefore, although high pH soils were more diverse at a particular location (soil core), there was much less variability among groups of high pH soils. Overall, the mean total diversity (γ diversity $= \alpha + \beta$) remained lower in the low pH soils, but γ diversity increased less rapidly with increasing pH compared with α diversity (γ diversity $r^2 = 0.26$). One plausible mechanism for the decline in β diversity with increasing
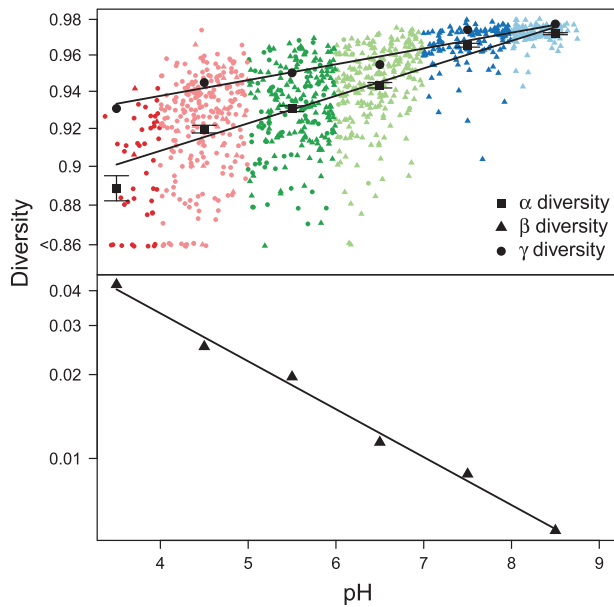
Fig. 1. Relationships between soil pH and bacterial diversity. The α diversity of soil bacterial communities was most strongly related to soil pH ($r^2 = 0.41$, $n = 1010$). While α diversity is positively related to pH, a negative relationship is observed for β diversity. Data points represent the Simpson's diversity index for each sample, with circles or triangles representing the aggregate vegetation classification (AVC): Circles = moorland grass mosaics, upland wooded, heath and bog; triangles = crops and weeds, fertile grassland, tall grass and herb, infertile grassland, lowland wooded habitats. Mean α, β and γ diversity calculated on soils grouped by pH are shown as indicated in the legend, with coloured data points representing the pH bins.

pH is that low pH soils are more variable across sites in their abiotic conditions.

We hypothesized that low pH organic soils represented relatively heterogeneous habitats, thereby facilitating a greater number of potential niches for bacterial communities compared with more homogeneous higher pH mineral soils. Variance across the pH gradient was examined by comparing the variation in abiotic conditions (climate, altitude, soil organic matter, soil moisture and soil phosphorous) within each of the pH classes (see Fig. 1). As predicted, there was a substantial decline in the variance in environmental conditions across sites (Fig. 2), lending support for the hypothesis that there are a greater number of potential niches in low pH soils.

Mapping and explaining variance in bacterial community composition

While diversity indices are useful in describing community characteristics by a single vector, they provide no evaluation of important compositional features of biodiversity relating to the abundances of shared taxa. To specifically assess changes in bacterial composition (incorporating taxon abundance and identity), we used non-metric mul-

tidimensional scaling (NMDS) ordination to represent, in two dimensions, the pairwise Bray Curtis similarities between soil community TRFLP profiles ($n = 1010$, Fig. 3A). Linear fitting of environmental factors to the ordination ($n = 818$) revealed significant effects of most measured variables, with pH again being the strongest correlate (Table 1). Consistent with higher β diversity, the spread of data points over the ordination was greater in low pH soils. Therefore, the higher pH soils, while comprising a larger diversity of taxa, are more similar in terms of community distances than the less diverse low pH soils.

The significance of the other measured variables can be interpreted by examining the directions of the correlations with the samples scores in the ordination plot (Table 1) or by mapping the ordination scores across Great Britain (Fig. 3B). Northern regions posses a greater proportion of upland habitats, with the harsh climatic conditions giving rise to distinct plant communities and acidic soils of high organic matter content. These factors show commonality with heathland and upland regions of England and Wales, and thus these biomes shared similar bacterial communities. This is in marked contrast to most of England and parts of Wales, which generally possess a milder climate, less acidic lowland soils and a typically agricultural floristic community (Haines-Young *et al.*, 2003). These areas share common characteristics with lowland regions of Eastern Scotland, and predictably, their bacterial communities were more similar. As exemplified by the high correlations with plant community ordination
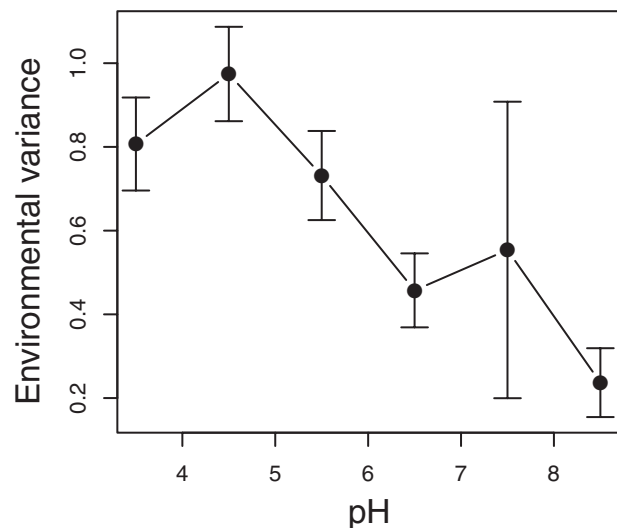

Fig. 2. Relationship between the variance in environmental conditions and soil pH. Each of the quantitative abiotic environmental measurements was scaled to have a mean of zero and standard deviation of one. For each environmental parameter, the variance was calculated over each of the pH bins used in Fig. 1. The mean (+/− standard error) of these variances is plotted across the pH range.
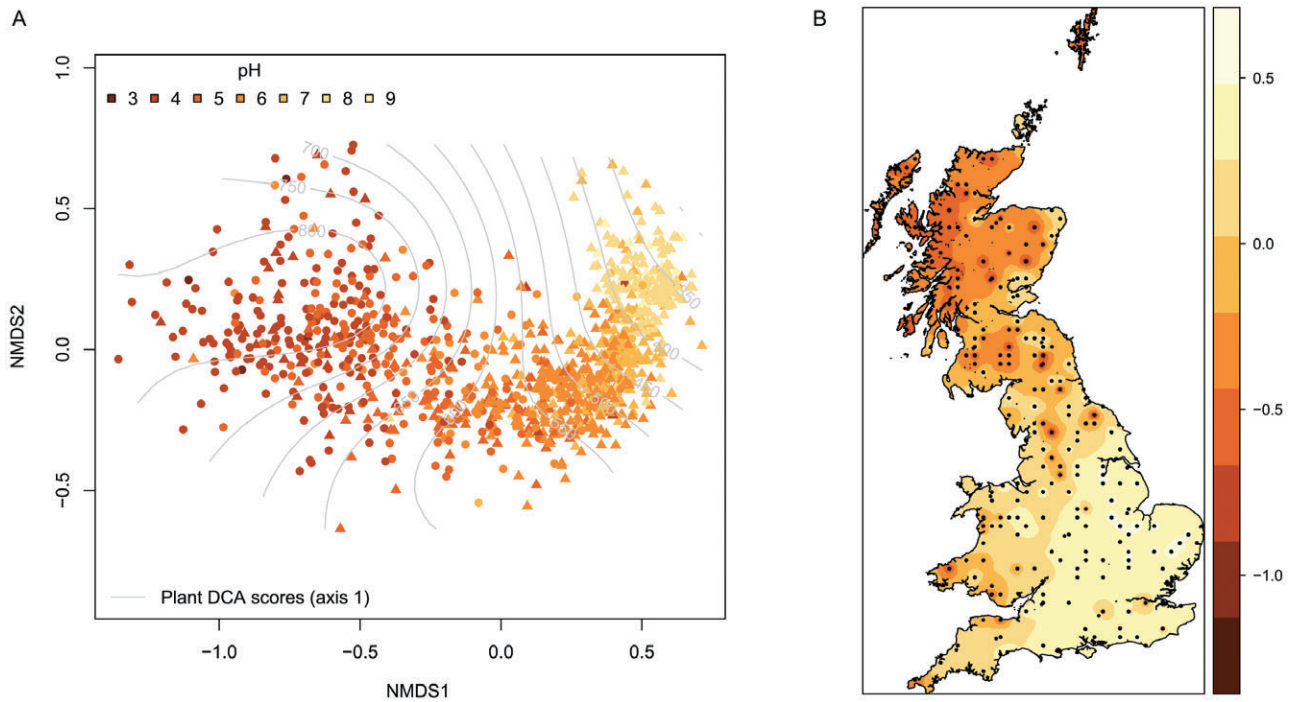
Fig. 3. Ordination and mapping of bacterial communities.
A. Representation of the pairwise similarities between samples using two dimensional NMDS ordination. The three dominant environmental gradients associated with the ordination are also represented: a colour ramp showing pH; circles or triangles representing the habitats defined by AVC group (see Fig. 1) and generalized additive model fitted contours illustrate the first axis sample scores obtained from detrended correspondence analyses of plant communities.
B. Spatial mapping of bacterial communities across Great Britain using inverse distance weight interpolation of first axis NMDS scores. The locations of the 1 km² plots that comprise up to five soil cores are indicated on the map. The main differences in bacterial communities are reflected in the broad environmental, climatic and land use features of the landscape.

Table 1. Summary of relationships between environmental factors and bacterial communities.

| Variable | r^2 | P | Correlation | |
			Axis1	Axis2
pH	0.777	0.001	+	+
Plant DCA scores (axis 1)	0.616	0.001	−	−
C : N Ratio	0.532	0.001	−	+
% Moisture	0.531	0.001	−	+
Loss on ignition (%)	0.506	0.001	−	+
% Carbon	0.495	0.001	−	+
% Nitrogen	0.405	0.001	−	+
Rain	0.349	0.001	−	−
Temperature (°C)	0.333	0.001	+	+
Sun (h)	0.327	0.001	+	+
Eastings (m)	0.321	0.001	+	+
Cloud (% cover)	0.304	0.001	−	−
Altitude (m)	0.261	0.001	−	−
Plant DCA scores (axis 2)	0.242	0.001	−	−
Northings (m)	0.195	0.001	−	−
Phosphorus (mg kg⁻¹)	0.046	0.001	+	+
Plant DCA scores (axis 3)	0.002	0.5	−	+

Correlations with numeric environmental variables (r^2) were obtained by fitting linear trends to the NMDS ordination and significance (P) was determined by permutation (nperm = 999). +/− signifies the direction of association with each of the NMDS axes.

results (Axis 1 of a detrended correspondence analysis: See Table 1 and Fig. 3A), soil bacterial communities are broadly structured similarly to plant communities with distributions governed by covarying soil chemical, geological, climatic conditions together with historical and current land use.

The role of space and environment

Here we wish to explore whether there are any additional spatial sources of variation in bacterial communities indicative of either local biotic processes such as dispersal, or other unmeasured spatially correlated environmental parameters. We address this question in terms of distances by examining the correlations between community, environmental and spatial distance matrices using Mantel tests (Legendre and Legendre, 1998). We maximized the community–environment correlation by selecting a subset of environmental parameters giving the highest Pearson's *r*. This calculation was performed using the bioenv procedure within the R package vegan (Oksanen *et al.*, 2009) and resulted in the selection of three variables [pH, plant DCA 1 scores and (log)CN ratio], which were then centred on their means and scaled
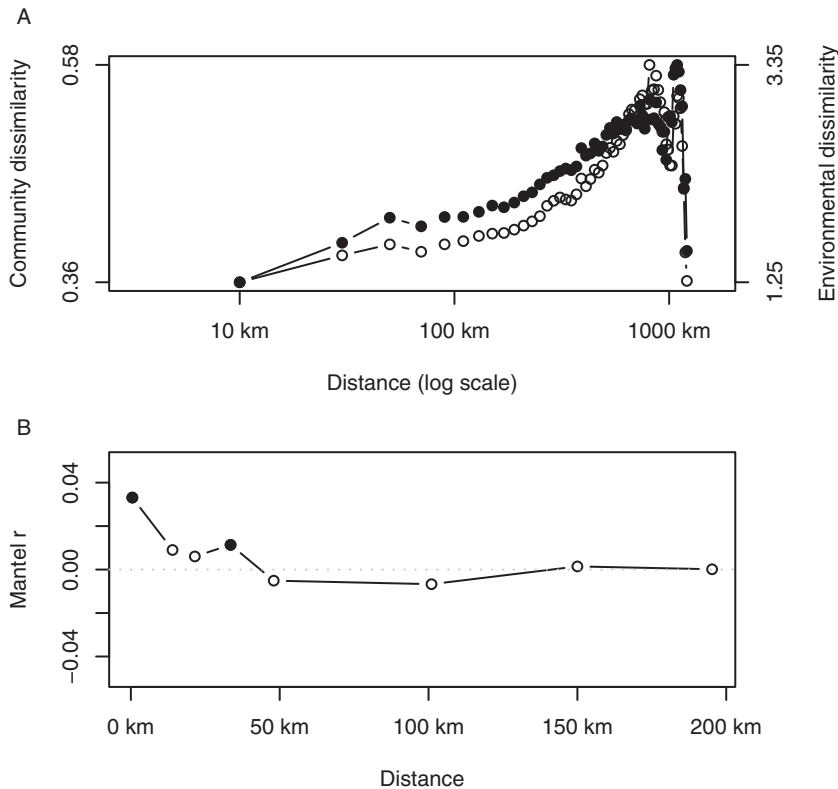
A

B

Fig. 4. Relationships between community, spatial and environmental distances.
A. The mean variance in community (filled circles) and environmental (open circles) dissimilarities with space. Means were calculated over discrete distance lags of 20 km length across all sampling sites, and the data rescaled to facilitate plotting. Both community and environmental dissimilarities increase with distance, a consequence of the high level of environmental control over bacterial communities.
B. Mantel correlogram showing spatial autocorrelation of the community residuals after linear regression with environmental distances. The first distance lag was selected to include all comparisons at < 1 km and subsequent lags chosen to represent approximately equal *n* representative of distances up to 200 km. The significance of each correlation (filled circles) with all other lags was tested by permutation ($n = 999$, $\alpha < = 0.001$).

to unit variance before calculation of the environmental distance matrix. When examining the total dataset ($n = 886$ samples after removing missing environmental data, 392 055 pairwise comparisons), mean community and environmental dissimilarity increased in similar manner with geographic distance (Fig. 4A). However, a thorough evaluation of partial residual plots revealed non-linear patterns within the datasets, potentially invalidating the results of Mantel tests. These are inferred in Fig. 4A, particularly at the furthest distance classes, where environmental dissimilarity continues to rise whereas community dissimilarity levels off. Additionally at the most extreme distance lags relationships become erratic, as a result of a combination of a lower number of pairwise comparisons, and broadly similar biomes at the most northerly and southerly locations within Britain. These issues were overcome by truncating the distance matrices to only include comparisons between samples less than 200 km apart, and then taking the logarithm of these distance before the Mantel tests.

Both community and spatial dissimilarities were positively correlated with geographic distance (Mantel $r = 0.10$ and 0.11, respectively, $P < 0.001$). However, the community–space relationship was marginal compared with the correlations of community and environmental dissimilarities (Mantel $r = 0.6973$, $P < 0.001$). Partial correlations of community distances and space, given the environment, resulted in a weak but significant correlation (partial Mantel $r = 0.035$, $P < 0.001$), indicating a minor increase in genetic dissimilarity with space once the environmental relationships had been removed. Further evidence of a significant spatial component for the closest distance classes was observed by constructing a partial mantel correlogram (Fig. 4B), which tests the mean residual community variation (after environment) for discrete distance classes against the means over all other distances (Legendre and Legendre, 1998). Here significant correlations were obtained for the first distance class (< 1 km), indicating that the mean residual variance within the 1 km squares was less than between square variation once environment had been taken into account. However, this spatially correlated residual component was still marginal compared with the large correlations between community and environment. It is of note that other methods based on constrained ordination and variance partitioning (Legendre *et al.*, 2008) were also explored to estimate the relative importance of environmental and spatial factors. While these methods also highlighted the strong importance of environmental structuring, the appropriateness of their usage was uncertain as the ordinations were notably arched, indicative of non-linear correlations present over long environmental gradients. We therefore conclude that most of the variation in soil bacterial communities determined with the TRFLP method can be explained by the

environment, although there is some evidence to suggest other variables not assessed here may additionally structure communities at local scales.

Broad taxonomic features of soil bacterial communities

The identities of bacterial taxa differing across the environmental gradients were investigated by sequencing a subset of samples. Approximately 160 16S rRNA gene clones were sequenced from each nucleic acid extract from 15 geographically dispersed soils of low, medium and high pH ($n = 5$; ~400 nucleotides per clone). Sequences were then grouped and classified into operational taxonomic units (OTUs) at varying levels of similarity, and the results confirmed patterns shown by the TRFLP method. Regardless of the similarity level used to classify sequences, low pH soils had less taxonomic richness and diversity compared with intermediate or high pH soils (Fig. S1A and B). However, maximal correlations between TRFLP and clone library diversity indices ($r^2 \sim 0.8$) were found when sequences were classified at a low level of sequence similarity (60–70%) implying that the TRFLP approach does not discriminate taxa at a particularly high taxonomic resolution (Fig. S2). Assessments of the numbers of distinct taxa found in all five soils from each pH group also highlighted that the compositional distinctness of equivalent pH soils is only manifest when sequences are compared at a relatively low taxonomic resolution (Fig. S1C). Taxa could only be found consistently in all five soils of similar pH when sequences were classified at low levels of similarity. Conversely, when grouping sequences at the 0–10% dissimilarity ranges, there were in fact few taxa common to all five soils of equivalent pH. We therefore present only the broad taxonomic groups that differed between samples across the pH gradient (Fig. 5), following classification using the Ribosomal Database Project's Bayesian Classifier. Low pH soils were dominated by Group 1 Acidobacteria along with the Alphaproteobacteria. As pH increased the Alphaproteobacteria became more dominant at intermediate pH, and higher pH soils comprised a more evenly distributed abundance of many different taxa, with a notable increase in the Actinobacteria and Group 6 Acidobacteria.

In order to consolidate the sequencing data and TRFLP community profiles, we performed an exploratory analysis to determine the sequence identity of dominant TRFLP peaks responding to the dominant environmental gradient. We first examined individual terminal restriction fragments (TRFs), which had strong associations with the NMDS axis scores (in Fig. 3A), and then determined the identities of these fragments by examining enzyme cut sites within the available sequences (Fig. S3A). Communities to the left of the NMDS ordination plot were domi-

nated by several peaks, which could be confidently identified as Group 1 Acidobacteria (52 to 55 and 226 to 229 nucleotides). Plotting the relative abundances of these TRFs versus pH revealed a marked decline in Group 1 Acidobacteria with an increase in pH across all samples analysed in the survey (Fig. S3B). With increasing first axis sample scores (reflecting a positive pH gradient) the negative skew on the second axis corresponded with an increased abundance of several TRFs identifiable as Alphaproteobacteria, which exhibited a unimodal response to pH. The high pH soils to the right of the ordination comprised a high diversity of many different taxa. Both the TRFLP and sequencing approaches therefore revealed that the strong environmental effects on bacterial communities were largely due to the responses of only a few broad taxonomic lineages to the environmental gradients.

Discussion

The results from this study show that bacterial diversity measured with simple community profiling methodologies is mainly correlated with soil pH, in broad agreement with other studies examining soil bacterial biogeography across America (Fierer and Jackson, 2006; Lauber *et al.*, 2009). However, the exact shape of the relationship differed within British soils despite this survey covering similar ranges of soil pH. No decline in diversity at high pH was observed in our study, possibly identifying unexplored interactions between pH and diversity within the southern dry habitats examined in the American survey. Similarly, other more recent studies investigating local relationships between pH and diversity in temperate and Arctic biomes have failed to identify reduced diversity at high pH (Hartman *et al.*, 2008b; Rousk *et al.*, 2009; Chu *et al.*, 2010), illustrating that more widespread surveys are required to provide a global synthesis on the exact nature of this relationship. Importantly, we also found that other measured environmental factors also had a significant influence on soil bacterial diversity and community dissimilarities. This could be due to subtle differences in methodology, or the differences in sampling scale and location carried out here. It is possible that the tenfold higher sample sizes in the current study provided the statistical power required to detect more subtle environmental drivers of bacterial biodiversity.

The large correlations between bacterial and plant biodiversity observed in our study, albeit confound by soil pH, reveals that there are direct correlative links between above and belowground diversity. The biogeography of dominant broad groups of bacterial taxa is therefore determined by covarying soil chemical, geological, climatic and biotic factors commonly known to affect the distribution of larger organisms (Gaston, 2000). Addition-
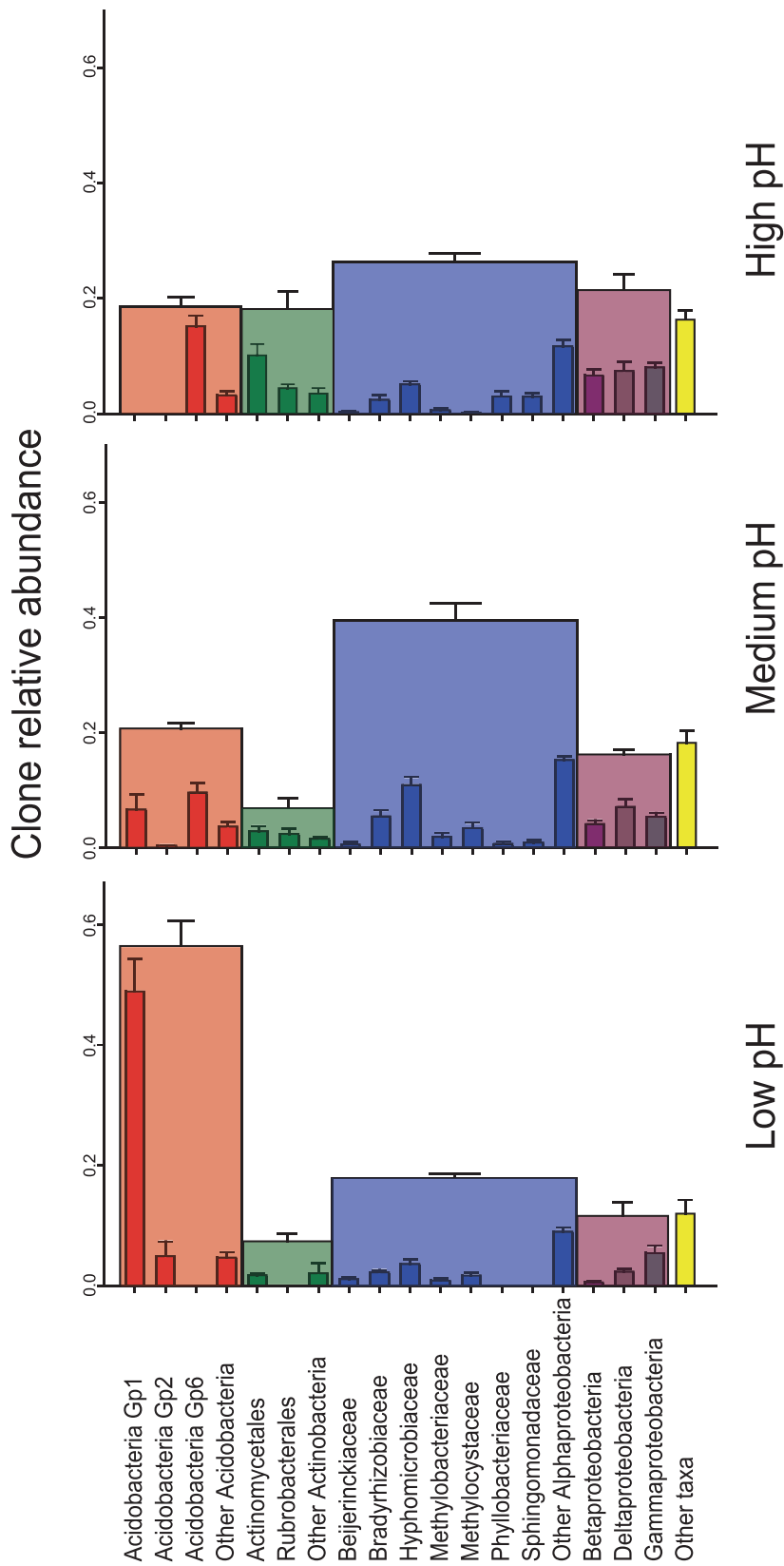
Fig. 5. Relative abundances of taxa at low, medium and high pH. Clone libraries were sequenced from a range of geographically dispersed soils at low ($n = 5$, 788 total sequences), medium ($n = 5$, 879 total sequences) and high ($n = 5$, 889 total sequences) pH. The mean relative abundances (+/− standard error) of the dominant detected phyla are shown together with internal bars showing the abundances at the next taxonomic subdivision. Colours of bars indicate broad phylum level affiliation: Acidobacteria (red), Actinobacteria (green), Alphaproteobacteria (blue), other Proteobacteria (purple) and other taxa (yellow).

ally, we infer that alterations in land use that affect soil pH may have large consequences for soil bacterial biodiversity at landscape scales. Establishing whether pH alone is directly responsible for the observed differences in composition is problematic as soil pH is correlated with a number of other biotic and abiotic variables. Soil pH affects C and N substrate availability, possibly brought about through a change in aboveground communities (Kemmitt *et al.*, 2006); and increased bioavailability of toxic metals is also noted in low pH soils (Pietri and Brookes, 2008), which may have additional effects upon soil microbes.

An important aspect of our survey is that it represents the first large-scale molecular examination of soil bacterial 16S rRNA gene β diversity. For microbial communities the variation in diversity between samples (β diversity) has traditionally been less studied than α diversity (diversity within a single sample), despite β diversity measurements being considered important for an overall understanding of community dynamics (Green and Bohannan, 2006; Lozupone *et al.*, 2007). Contrary to α diversity, β diversity was highest in low pH, high C organic soils and lowest in high pH, low C agricultural soils. Similarly, in marine sediments β diversity was shown to be higher in low pH, high C surface layers, and decreased with increasing depth and pH and decreasing C concentrations (Wang *et al.*, 2008). In British soils, decrease in β diversity correlated with a reduction in environmental heterogeneity among soils across the pH gradient. Other alternative, yet complementary, explanations may relate to greater vertical heterogeneity in low pH soil cores, and also the possibility that less diverse communities are naturally more variable and prone to turnover. Hypothetically, multiple soil cores taken across a bog are likely to differ more in their abiotic and biotic characteristics than multiple cores taken across an improved grassland field. The limited number of broad taxonomic groups found in the acidic soils will therefore vary more in relative abundance across the bog, possibly because of a greater number of possible niches within the environmentally patchy landscape. We did not find evidence in our study that greater β diversity meant meta-communities of low pH soils matched or exceeded the total diversity (γ diversity) of higher pH soils. However, these results emphasize the importance of considering between-site variance in future studies attempting to form landscape scale conclusions regarding diversity from point sample assessments.

The differences in biodiversity across the detected environmental gradients were the result of consistent changes in the composition of the dominant broad taxa, identified by both the TRFLP analyses and the sequencing of a limited number of soils. The dominant groups of bacteria detected (Fig. 5) are known to be the most abundant in

soils globally (Janssen, 2006), and we observed similar responses of these taxa to soil pH as reported in another study, which used pyrosequencing (Lauber *et al.*, 2009). The Alphaproteobacteria and Actinobacterial groups have been known for some time to play important roles in soil, a facet of the general culturability of these taxa (Hugenholtz *et al.*, 1998). Additionally, there is growing recognition of the importance of Acidobacteria in the soil environment (Barns *et al.*, 1999; Jones *et al.*, 2009). Previously a broad lineage of taxa only discovered in soils through the use of molecular techniques, they have now been shown to be culturable using low pH media, low nutrients and long incubation times (Kishimoto *et al.*, 1991; Barns *et al.*, 1999; Sait *et al.*, 2002; 2006).

Our findings support other molecular studies that reveal that the Acidobacteria phylum, in general, are more abundant at low pH, yet display subgroup differences in response to soil pH: Groups 1 and 2 Acidobacteria were negatively correlated with soil pH, whereas Group 6 Acidobacteria increased in abundance with increasing pH (Lauber *et al.*, 2008; 2009; Jones *et al.*, 2009; Rousk *et al.*, 2010). Importantly, the low α diversity and high β diversity at low pH were due to the large, yet variable, abundances of the Group 1 Acidobacteria. Therefore, it appears that the established pH–diversity relationship observed here and elsewhere is driven by the dominance of a few taxonomic groups in low pH soils. It remains to be determined whether the pH effect is still apparent when rarer members of the community are assessed; with these particular groups excluded. Indeed this represents a potential question, which could be addressed by future targeted deep sequencing. Furthermore, it must be noted that importance of pH in driving changes in broad taxa does not negate the influence of other environmental, temporal and biotic processes in driving the diversity of sub-populations of these taxa. However, the changes in dominant taxa over the pH gradient may have consequences for soil processes as there is increasing evidence that Alphaproteobacteria (the other predominant taxon at low pH) dominate Acidobacteria in soils with high nutrient availability, and the relative abundances of these taxa have been associated with broad measures of ecosystem functioning such as carbon cycling (Smit *et al.*, 2001; Griffiths *et al.*, 2006; Fierer *et al.*, 2007; Thomson *et al.*, 2010). At the simplest level, it is imperative that we understand the functional consequences of these broad shifts in dominating taxonomic lineages over soil environmental gradients, if we are to progress in deciphering exactly how bacterial communities are linked with soil processes. Additionally, if such linkages do in fact exist, quantifying the variance of dominating lineages in similar soil biomes (β diversity) may be useful in explaining the variability of soil processes at landscape scales.

Despite the large effect of environmental factors on community distances, we also found some evidence for residual spatial autocorrelation at closer spatial scales. However, given the relatively low taxonomic resolution of the TRFLP method identified subsequently in this study and elsewhere (Blackwood *et al.*, 2007), we do not believe this is a significant indicator of structuring as a result of biotic factors such as dispersal. More likely, it is indicative of other spatially correlated environmental factors not included in the analyses. One such parameter that was not assessed in our study was the specific time each core was sampled during the 2007 summer, which could reflect local variability in climate or plant mediated nutrient cycling, factors that are known to influence bacterial communities (Griffiths *et al.*, 2003). Regardless of these minor effects, the general dominance of control by edaphic factors is in contrast to several studies that have used methods discriminating taxa at higher taxonomic resolution to explore environment–space relationships (Cho and Tiedje, 2000; Ramette and Tiedje, 2007). Similarly, a study using pyrosequencing to contrast bacterial biodiversity in different soils found little taxonomic overlap across sites (Fulthorpe *et al.*, 2008). This finding in particular was mirrored in our clone analyses, which showed few shared taxa in distantly sampled soils of equal pH, when OTUs were defined at low levels of sequence dissimilarity.

Therefore, the larger effect of edaphic control is only manifest when assessing sequences at the level of broad taxa, or using a simple community profiling tool with inherently low levels of taxonomic discrimination. It is likely that sub-populations of the taxonomic groups detected within these soils will be endemic at local scales, and we urge caution in confusing the strengths of environment–space relationships (or indeed any other ecological pattern) when communities are assessed at low levels of taxonomic resolution (for discussion see Horner-Devine *et al.*, 2004). In addition, it is noteworthy that seldom are such long environmental gradients assessed in community studies of higher organisms, and most multivariate statistical routines underperform when analysing such data (see Smith and Lundholm, 2010 for discussion). Future work investigating the biogeography of soil bacterial communities should therefore focus on local communities in soils of equivalent pH, using methods offering greater taxonomic discrimination. For example, an important question arising from our study is whether the magnitude of purely spatial variability relates to the changes in β diversity we observed along the soil pH gradient. Our data therefore serve to highlight the distinct soil biomes comprising similar broad community structures (see summary Fig. 6), which may provide a guide for future efforts specifically addressing local biogeographic patterns.

To conclude, this study represents one of the largest attempts to comprehensively map and investigate the regulation of different components of soil bacterial biodiversity at multiple spatial scales over a landscape. We have shown that broad groups of soil bacterial taxa are distributed non-randomly, and their abundances can be predicted by key environmental variables known to be important in structuring plant diversity. Of fundamental importance, we believe that neither our study nor other studies assessing broad taxonomic groups of bacteria provide sufficient taxonomic discrimination to indicate that soil bacterial biogeographic patterns are in any way unique. Future research in this area would benefit from using appropriate methodologies to discriminate bacterial taxa at a taxonomic resolution more reflective of the levels used in generating theory for larger organisms (Horner-Devine *et al.*, 2004; Martiny *et al.*, 2006). Meanwhile, the emerging global recognition of the dominant broad taxa inhabiting specific soil biomes should be marked as an achievement in itself, and it is now imperative to address the significance of these compositional differences within a functional context. Determining whether these basic taxonomic differences confer altered functionality will enable a true evaluation of the role soil bacterial biodiversity in modulating ecosystem services, and provide a framework to predict landscape scale responses under future environmental change scenarios.

Experimental procedures

Sampling regime

Samples were collected between May and November 2007 from 233 1 km^2 squares across the UK as part of the Countryside Survey (http://www.countrysidesurvey.org.uk/). Within each 1 km^2 sampling area, up to five soil cores were sampled (5 cm diameter, 15 cm deep) from the centre of randomly allocated 200 m^2 sub-plots used for the vegetation survey. For each soil core the location (eastings, northings and altitude) was recorded along with field measures of flora. Floral assessments documented the presence and percent cover of vascular plants, and a selected list of the more common bryophytes and macro-lichens (Smart *et al.*, 2003). As predictor variables indicative of floral biodiversity we used vectors representing the first three axes of a Detrended Correspondence Analysis performed on the binary floral dataset (denoted DCA1, DCA2, DCA3). Additionally, a categorical variable was used to denote habitat type based upon the plant species assessments, denoted as the aggregate vegetation class (Firbank *et al.*, 2003). Soil physical and chemical characteristics were determined from a duplicate 15 cm core taken adjacent to the microbial core [for full details see: CS Technical report No. 3/07 (Emmett *et al.*, 2008)]. These numeric variables included pH; % carbon (C), % nitrogen (N), C : N Ratio, % organic matter (loss on ignition), phosphorous (Olsen P, mg kg^{-1}); and soil moisture content (% moisture).
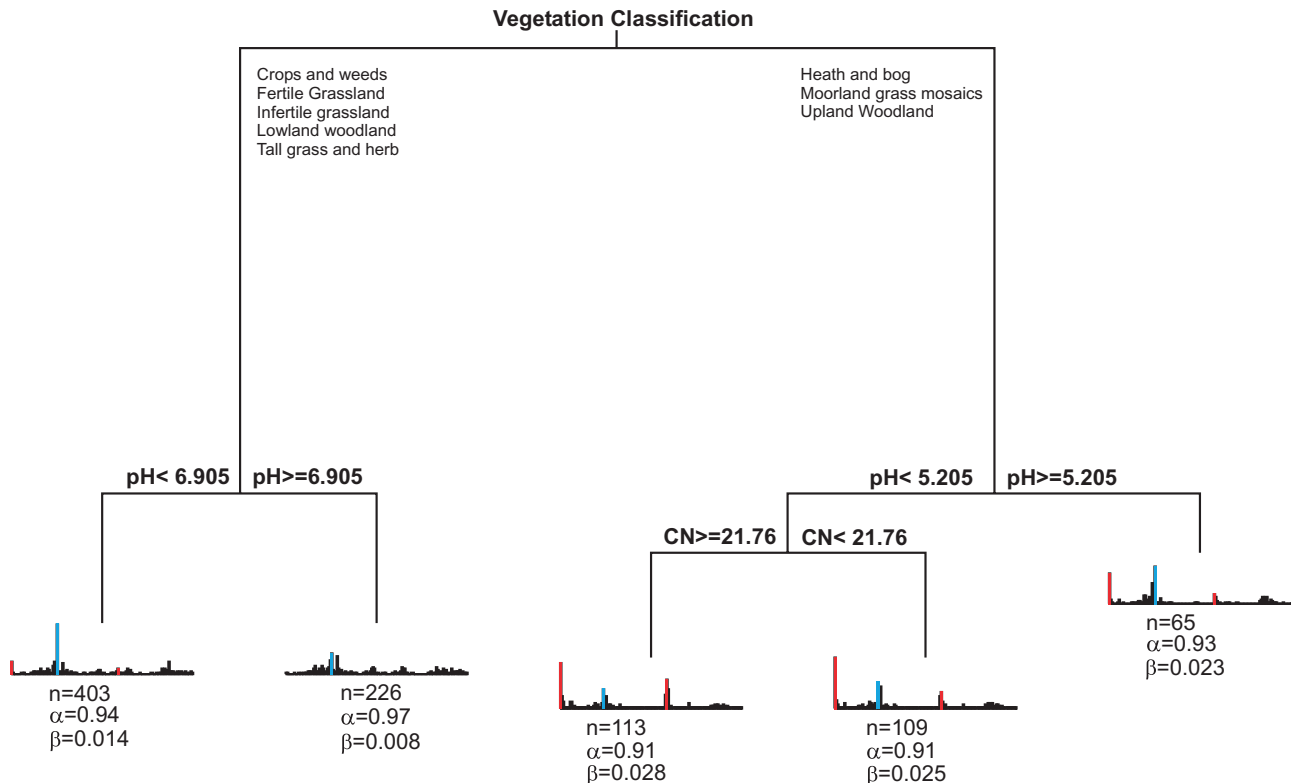
Fig. 6. Multivariate regression tree summarizing community–environment relationships. The tree was calculated on the TRFLP dataset using all the environmental variables as predictors. For each split a rule is selected based on the predictors to minimize the dissimilarity within the TRFLP profiles in the resulting two nodes. The tree explains 50% of the variability in TRFLP profiles, much of which is accounted for by the first split based on the vegetation classification. Soil pH was the next best predictor for the first split (pH < > 5.2), and also was the best predictor for the two subsequent splits. Low pH soils were additionally split into two groups based on soil C : N ratio. At each terminal node, the mean abundances of each TRF are shown, together with the mean α diversity and β diversity for each group. TRFs with an important influence on the community variance observed in this study are also highlighted (red = Group 1 Acidobacteria, blue = Alphaproteobacteria).

Climatic variables (rain, sun, cloud cover and temperature) were based on mean data per square between 1961 and 2000.

Molecular analyses

Soil cores were homogenized under sterile conditions and total nucleic acids were extracted from 0.25 g of soil using a previously described method (Griffiths *et al.*, 2000), modified to include a 30 min hexadecyltrimethylammonium bromide (CTAB) freeze-thaw, soft-lysis stage. TRFLP of 16S rRNA genes (V1–V3 variable regions) was performed using forward primer 63F 5′-CAGGCCTAACACATGCAAGTC-3′ labelled at the 5′ end with 6FAM fluorescent dye and reverse primer, 519R 5′-GTATTACCGCGGCTGCTG–3′ (Thomson *et al.*, 2010). Amplicons were purified using the PureLink PCR purification kit (Invitrogen, Paisley, UK), then digested using restriction endonuclease MspI. Fragment analysis was performed using a 3730 DNA analyser (Applied Biosystems, CA, USA) and individual TRFs were binned manually using Genemarker software (SoftGenetics, PA, USA). Before statistical analyses the intensity of each fragment was converted to a proportional abundance, by dividing by the total intensity of all detected fragments.

To identify individual TRFs changing across environmental gradients, a subset of 15 soil cores were selected for clone library generation and sequencing. These soils represented five spatially and environmentally distinct replicates at pH 4.23 (+/– 0.23), 6.15 (+/– 0.08) and 8.28 (+/– 0.16). Amplicons for cloning were generated from total DNA with primers 63F and 519R using the TOPO TA Cloning Kit (Invitrogen, Paisley, UK). Colonies were picked and colony PCR products (M13F and M13R primers) were sequenced using the M13R primer and BigDye v3.1 chemistry before analysis on a 3730 DNA analyser (Applied Biosystems, CA, USA).

Bacterial 16S rRNA gene sequences were trimmed and edited, before phylogenetic placement using the Ribosomal Database Project's Naive Bayesian rRNA Classifier tool (http://rdp.cme.msu.edu/classifier/) with a bootstrap cut-off of 50% (Wang *et al.*, 2007). Using the MOTHUR software (Schloss *et al.*, 2009) sequences were aligned to the greengenes 16S rRNA gene sequence database (DeSantis *et al.*, 2006) and an uncorrected pairwise distance matrix was formed. Sequences were then assigned to OTUs using the furthest neighbour clustering algorithm before calculation of diversity indices by classifying sequences at varying levels of similarity. Sequences were deposited in The European Molecular Biology Laboratory sequence database under

accession numbers FN808422-FN810977. To approximate the identities of individual TRFs, the obtained sequences were digested *in silico* with restriction endonuclease MspI using the TRFLPMAP software (http://nebc.nerc.ac.uk/cgi-bin/trflp0_2.cgi). Bacterial identifications were assigned to TRFs if greater than 75% of sequences for a given range of fragments belonged to a particular taxon.

Statistical analyses

All statistical analyses were carried out using various libraries within the R software package (R Core Development Team, 2005). Diversity indices were calculated by additively partitioning Simpson's diversity index according to Lande (Lande, 1996). The Simpson's index of diversity was first calculated for each sample (Eq. 1), where p_i represents the relative abundance of each TRFLP peak within each sample. Subsequently samples were binned into six discrete pH classes (pH 3.5 to 8.5) and the mean Simpson's index taken to represent the α diversity per pH class. β diversity was determined as a mean variance per pH class according to Eq. 2, where p_{ij} is the TRF abundance in each sample and $\bar{p}_i$ is the mean abundance of the ith TRF across all samples within the pH class. The γ diversity per pH class (Eq. 3) was taken as the sum of the mean α and mean β diversity.

Equation 1: α diversity $= 1 - \sum (p_i)^2$

Equation 2: β diversity $= \sum (p_{ij} - \bar{p}_i)^2$

Equation 3: γ diversity $= \alpha + \beta$

Non-metric multidimensional scaling analyses were performed using MetaMDS functions within the vegan package (Oksanen *et al.*, 2009) based on dissimilarities calculated using the Bray–Curtis index, and environmental vectors were fitted using the envfit and ordisurf routines. The variable scores (individual TRFs) associated with the NMDS ordination were calculated as weighted averages of site scores using the wascores function in vegan. First axes site scores were mapped using inverse distance weighting interpolation using the default settings within the gstat package (Pebesma, 2004). Mantel tests and mantel correlograms were carried out on Bray–Curtis distances for the species (TRFLP) data; euclidean distances for the geographical coordinates and scaled environmental variables within vegan. A sums of squares multivariate regression tree was calculated within the mvpart package, using the using the one-standard error rule on the cross-validated relative error to determine the number of terminal nodes (De'ath, 2002).

References

Amann, R., and Ludwig, W. (2000) Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. *FEMS Microbiol Rev* **24:** 555–565.

Baker, K.L., Langenheder, S., Nicol, G.W., Ricketts, D., Killham, K., Campbell, C.D., and Prosser, J.I. (2009) Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. *Soil Biol Biochem* **41:** 2292–2298.

Bardgett, R.D., Freeman, C., and Ostle, N.J. (2008) Microbial contributions to climate change through carbon cycle feedbacks. *ISME J* **2:** 805–814.

Barns, S.M., Takala, S.L., and Kuske, C.R. (1999) Wide distribution and diversity of members of the bacterial kingdom Acidobacterium in the environment. *Appl Environ Microbiol* **65:** 1731–1737.

Bent, S.J., Pierson, J.D., Forney, L.J., Danovaro, R., Luna, G.M., Dell'Anno, A., and Pietrangeli, B. (2007) Measuring Species Richness Based on Microbial Community Fingerprints: the Emperor Has No Clothes. *Appl Environ Microbiol* **73:** 2399–2401.

Blackwood, C.B., Hudleston, D., Zak, D.R., and Buyer, J.S. (2007) Interpreting ecological diversity indices applied to terminal restriction fragment length polymorphism data: insights from simulated microbial communities. *Appl Environ Microbiol* **73:** 5276–5283.

Brussaard, L. (1997) Biodiversity and Ecosystem Functioning in Soil. *Ambio* **26:** 563–570.

Cho, J.C., and Tiedje, J.M. (2000) Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. *Appl Environ Microbiol* **66:** 5448–5456.

Chu, H., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R., and Grogan, P. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ Microbiol* **12:** 2998–3006.

De'ath, G. (2002) Multivariate regression trees: a new technique for modeling species – environment relationships. *Ecology* **83:** 1105–1117.

Dequiedt, S., Thioulouse, J., Jolivet, C., Saby, N.P.A., Lelievre, M., Maron, P.-A., *et al.* (2009) Biogeographical patterns of soil bacterial communities. *Environ Microbiol Rep* **1:** 251–255.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069–5072.

Dunbar, J., Takala, S., Barns, S.M., Davis, J.A., and Kuske, C.R. (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl Environ Microbiol* **65:** 1662–1669.

Emmett, B.A., Frogbrook, Z.L., Chamberlain, P.M., Griffiths, R.I., Pickup, R., Poskitt, J., *et al.* (2008) Countryside Survey Technical Report No. 3/07: Soils Manual. Swindon, UK: NERC/Centre for Ecology & Hydrology, 180 pp.

Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of Soil Bacterial Communities. *Proc Natl Acad Sci USA* **103:** 626–631.

Fierer, N., Bradford, M.A., and Jackson, R.B. (2007) Toward an ecological classification of soil bacteria. *Ecology* **88:** 1354–1364.

Firbank, L.G., Barr, C.J., Bunce, R.G.H., Furse, M.T., Haines-Young, R., Hornung, M., *et al.* (2003) Assessing stock and change in land cover and biodiversity in GB: an introduction to Countryside Survey 2000. *J Environ Manage* **67:** 207–218.

Fulthorpe, R.R., Roesch, L.F.W., Riva, A., and Triplett, E.W. (2008) Distantly sampled soils carry few species in common. *ISME J* **2:** 901–910.

Gans, J., Wolinsky, M., and Dunbar, J. (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309:** 1387–1390.

Gaston, K.J. (2000) Global patterns in biodiversity. *Nature* **405:** 220–227.

Green, J., and Bohannan, B.J.M. (2006) Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21:** 501–507.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., and Bailey, M.J. (2000) Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* **66:** 5488–5491.

Griffiths, R.I., Whiteley, A.S., O'Donnell, A.G., and Bailey, M.J. (2003) Influence of depth and sampling time on bacterial community structure in an upland grassland soil. *FEMS Microbiol Ecol* **43:** 35–43.

Griffiths, R.I., Bailey, M.J., McNamara, N.P., and Whiteley, A.S. (2006) The functions and components of the Sourhope soil microbiota. *Appl Soil Ecol* **33:** 114–126.

Haines-Young, R., Barr, C.J., Firbank, L.G., Furse, M., Howard, D.C., McGowan, G., *et al.* (2003) Changing landscapes, habitats and vegetation diversity across Great Britain. *J Environ Manage* **67:** 267–281.

Hartman, W.H., Richardson, C.J., Vilgalys, R., and Bruland, G.L. (2008a) Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proc Natl Acad Sci USA* **105:** 17842–17847.

Hartman, W.H., Richardson, C.J., Vilgalys, R., and Bruland, G.L. (2008b) Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proc Natl Acad Sci USA* **105:** 17842–17847.

Horner-Devine, M.C., Carney, K.M., and Bohannan, B.J.M. (2004) An ecological perspective on bacterial biodiversity. *Proc Biol Sci* **271:** 113–122.

Hugenholtz, P., Goebel, B.M., and Pace, N.R. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180:** 4765–4774.

Janssen, P.H. (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol* **72:** 1719–1728.

Jesus, E.D., Marsh, T.L., Tiedje, J.M., and Moreira, F.M.D. (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J* **3:** 1004–1011.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3:** 442–453.

Kemmitt, S.J., Wright, D., Goulding, K.W.T., and Jones, D.L. (2006) pH regulation of carbon and nitrogen dynamics in two agricultural soils. *Soil Biol Biochem* **38:** 898–911.

Kishimoto, N., Kosako, Y., and Tano, T. (1991) Acidobacterium-Capsulatum Gen-Nov, Sp-Nov – an Acidophilic Chemoorganotrophic Bacterium Containing Menaquinone from Acidic Mineral Environment. *Curr Microbiol* **22:** 1–7.

Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. *Oikos* **76:** 5–13.

Lauber, C.L., Strickland, M.S., Bradford, M.A., and Fierer, N. (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol Biochem* **40:** 2407–2415.

Lauber, C.L., Hamady, M., Knight, R., and 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.

Legendre, P., and Legendre, L. (1998) *Numerical Ecology*, 2nd edn. Amsterdam, The Netherlands: Elsevier Science.

Legendre, P., Borcard, D., and Peres-Neto, P.R. (2008) Analyzing or explaining beta diversity? Comment. *Ecology* **89:** 3238–3244.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007) Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73:** 1576–1585.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., O'Hara, R.G., Simpson, G.L., *et al.* (2009) Vegan: community ecology package. R package version 1.16-33 [WWW document]. URL http://R-Forge.R-project.org/projects/vegan/.

Pebesma, E.J. (2004) Multivariable geostatistics in S: the gstat package. *Comput Geosci* **30:** 683–691.

Pietri, J.C.A., and Brookes, P.C. (2008) Relationships between soil pH and microbial properties in a UK arable soil. *Soil Biol Biochem* **40:** 1856–1861.

Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al.* (2007) Essay – the role of ecological theory in microbial ecology. *Nat Rev Microbiol* **5:** 384–392.

R Core Development Team (2005) R: a language and environment for statistical computing. *R Foundation for Statistical Computing.* Vienna, Austria. ISBN 3-900051-07-0 [WWW document]. URL http://www.R-project.org.

Ramette, A., and Tiedje, J.M. (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci USA* **104:** 2761–2766.

Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., *et al.* (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* **1:** 283–290.

Rousk, J., Brookes, P.C., and Baath, E. (2009) Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Appl Environ Microbiol* **75:** 1589–1596.

Rousk, J., Baath, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G., *et al.* (2010) Soil bacterial and fungal

communities across a pH gradient in an arable soil. *ISME J* **4:** 1340–1351.

Sait, M., Hugenholtz, P., and Janssen, P.H. (2002) Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ Microbiol* **4:** 654–666.

Sait, M., Davis, K.E.R., and Janssen, P.H. (2006) Effect of pH on isolation and distribution of members of subdivision 1 of the phylum Acidobacteria occurring in soil. *Appl Environ Microbiol* **72:** 1852–1857.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Smart, S.M., Clarke, R.T., van de Poll, H.M., Robertson, E.J., Shield, E.R., Bunce, R.G.H., and Maskell, L.C. (2003) National-scale vegetation change across Britain; an analysis of sample-based surveillance data from the Countryside Surveys of 1990 and 1998. *J Environ Manage* **67:** 239–254.

Smit, E., Leeflang, P., Gommans, S., van den Broek, J., van Mil, S., and Wernars, K. (2001) Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl Environ Microbiol* **67:** 2284–2291.

Smith, T.W., and Lundholm, J.T. (2010) Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography* **33:** 648–655.

Thomson, B.C., Ostle, N., McNamara, N., Bailey, M.J., Whiteley, A.S., and Griffiths, R.I. (2010) Vegetation affects the relative abundances of dominant soil bacterial taxa and soil respiration rates in an upland grassland soil. *Microbiol Ecol* **59:** 335–343.

Wang, J.J., Wu, Y.C., Jiang, H.C., Li, C.H., Dong, H.L., Wu, Q.L., *et al.* (2008) High beta diversity of bacteria in the shallow terrestrial subsurface. *Environ Microbiol* **10:** 2537–2549.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261–5267.

Whittaker, R.H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol Monogr* **30:** 279–338.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Analyses of sequences and relationships with OTU definition. Sequences from 15 geographically dispersed soils at low, medium and high pH were classified into OTUs at incrementally varying levels of dissimilarity ($n = 5$ for each pH group). (A) Mean (+/− standard error) taxonomic richness, defined as the total numbers of OTUs, for low medium and high pH soils. (b) Mean (+/− standard error) Simpson's diversity for low medium and high pH soils. (C) The proportions of common sequences within each pH group. The numbers of sequences present in all five clone libraries at low, medium and high soil pH was quantified. Soils of similar pH share similar taxa only when sequences are defined at high levels of dissimilarity.

Fig. S2. Correlations of clone library and TRFLP data. The Simpson's diversity index was calculated for each of the clone libraries after classification of sequences at varying levels of similarity. The plot shows the relationship with the corresponding indices obtained from the TRFLP analyses of the same subset of soils. Strongest correlations are observed when clones are clustered at low levels of similarity, showing that the TRFLP method is unlikely to resolve closely related taxa.

Fig. S3. Responses of dominant taxa to environmental gradients (TRFLP data). (A) Plots showing the strength of associations between bacterial taxa (TRFLP peaks) and each of the NMDS axes. Most of the TRFLP peaks correlated with the negative first axis scores can be identified as Group 1 Acidobacteria (highlighted in red). Positive first axis scores are associated with a greater richness of TRFLP peaks. Most of the TRFLP peaks correlated with the negative second axis scores are identifiable as Alphaproteobacteria (shown in blue). (B) The relative abundances of these taxa plotted against the dominant environmental gradient (pH, $n = 1010$). High abundances of Group 1 Acidobacteria are associated with low pH-organic soils. The negative skew on the second axis is largely the result of a unimodal relationship between Alphaproteobacteria across the dominant environmental gradients. TRFLP profiles to the right of the ordination (high pH mineral soils) comprise a large number of unidentified taxa.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.