Environmental Microbiology (2008) **10**(10), 2537–2549 doi:10.1111/j.1462-2920.2008.01678.x

High beta diversity of bacteria in the shallow terrestrial subsurface

Jianjun Wang,[1,2] Yucheng Wu,[1] Hongchen Jiang,[3] Chunhai Li,[1] Hailiang Dong,[3] Qinglong Wu,[1] Janne Soininen[4] and Ji Shen[1]*

[1]*State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, East Beijing Road 73, Nanjing 210008, China.*

[2]*Graduate School of Chinese Academy of Sciences, Beijing, 100039, China.*

[3]*Department of Geology, Miami University, Oxford, Ohio 45056, USA.*

[4]*Department of Biological and Environmental Sciences, PO Box 65, FIN-00014 University of Helsinki, Finland.*

Summary

While there have been a vast number of studies on bacterial alpha diversity in the shallow terrestrial subsurface, beta diversity – how the bacterial community composition changes with spatial distance – has received surprisingly limited attention. Here, bacterial beta diversity and its controlling factors are investigated by denaturing gradient gel electrophoresis and cloning of samples from a 700-cm-long sediment core, the lower half of which consisted of marine-originated sediments. According to canonical correspondence analysis with variation partitioning, contemporary environmental variables explain beta diversity in a greater proportion than depth. However, we also found that community similarity decayed significantly with spatial distance and the slopes of the distance–decay relationships are relatively high. The high beta diversity indicates that the bacterial distribution patterns are not only controlled by contemporary environments, but also related to historical events, that is, dispersal or depositional history. This is highlighted by the different beta diversity patterns among studied sediment layers. We thus conclude that the high beta diversity in the shallow terrestrial subsurface is a trade-off between historical events and environmental heterogeneity. Furthermore, we suggest that the high beta diversity of bacteria is likely to be recapitulated in other terrestrial sites because of the great frequency of high geochemical and/or historical variations along depth.

Received 04 December, 2007; accepted 27 April, 2008. *For correspondence. E-mail Jishen@niglas.ac.cn; Tel. (+86) 25 8688 2005; Fax (+86) 25 5771 3063.

Introduction

The description of microbial diversity and its variation, or the assessment of the factors structuring the community composition and its variation, is of considerable interest to environmental microbiologists. Typically, there are two community parameters to characterize microbial diversity: alpha diversity (α) and beta diversity (β). Beta diversity, i.e. the variation of species composition along space or time, is a measure of difference in microbial community composition between pairwise sites. The pioneering investigations on beta diversity can be dated back to Whittaker (1972). However, only recently has beta diversity been explicitly and thoroughly examined for microorganisms within terrestrial or aquatic ecosystems (e.g. Green *et al*., 2004; Green and Bohannan, 2006; Lozupone *et al*., 2007; Shade *et al*., 2008), and further been used in a framework of microbial biogeography to examine the relative importance of habitat (contemporary environmental factors) and province (historical legacies) (Martiny *et al*., 2006; Takacs-Vesbach *et al*., 2008). In fact, microbial beta diversity is not less important than alpha diversity because information on beta diversity is expected to help in understanding the processes shaping microbial distribution pattern (Martiny *et al*., 2006), in designing systems for preservation of biodiversity (Green and Bohannan, 2006; Franklin and Mills, 2007), in managing microbial communities for bioremediation and even with developing ecological theories that can be applied to microorganisms (Hubbell, 2001; Prosser *et al*., 2007; Ramette and Tiedje, 2007a).

Microbes in the shallow terrestrial subsurface comprise an enormous amount of the Earth's biomass and species diversity (Whitman *et al*., 1998), and mediate important biogeochemical processes, including greenhouse gas emissions, organic matter mineralization, nitrogen cycling and transformation of pollutants (Holden and Fierer, 2005). Advances in understanding microbial communities (e.g. Zhou *et al*., 2002; Fierer *et al*., 2003; LaMontagne *et al*., 2003; Holden and Fierer, 2005; Allen *et al*., 2007; Barns *et al*., 2007; Hansel *et al*., 2008) have contributed to the understanding of microbial diversity in shallow terrestrial subsurfaces. In the first few meters of terrestrial

subsurface, microbial biomass, alpha diversity and activity often exhibit parallel declines with depth during the transition from organic-rich topsoil to mineral-dominated subsoil (e.g. Fierer *et al*., 2003; LaMontagne *et al*., 2003; Agnelli *et al*., 2004; Holden and Fierer, 2005; Lehman, 2007). These patterns imply a high degree of microbial turnover along environmental gradients or depth. However, our knowledge of the vertical distribution patterns of microbes in the shallow terrestrial subsurface is still very limited, and the degree of beta diversity and the controlling factors that structure microbial communities have been less rigorously quantified. Furthermore, the roles of environmental heterogeneity and spatial distance (a proxy for past historical events and disturbances) (Martiny *et al*., 2006) and their relative importance in shaping beta diversity in the shallow terrestrial subsurface still remain unexplored.

Both difficulties in sampling from the subsurface environment (Lehman, 2007) and the incomplete taxonomic definition of microbial species (Green and Bohannan, 2006; Martiny *et al*., 2006; Woodcock *et al*., 2006; Prosser *et al*., 2007) still hinder the mapping of microbial beta diversity and impede our understanding of its variation. One feasible method for addressing these problems is the application of molecular fingerprint methods, such as terminal restriction fragment length polymorphisms (LaMontagne *et al*., 2003) or DGGE (denaturing gradient gel electrophoresis) (van der Gucht *et al*., 2007), combined with multivariate statistical methods (Ramette, 2007). Despite some limitations, such as low resolution, fingerprint methods can be very useful in microbial biodiversity studies (Green *et al*., 2004; Fry *et al*., 2006; Woodcock *et al*., 2006; Marzorati *et al*., 2008). For example, as a well-established fingerprint method with a high versatility, reliability and reproducibility, significant shifts in DGGE patterns can be used as a proxy for real shifts in dominant bacterial community composition, both for comparative analysis of relative positions and abundances (Muyzer *et al*., 1993; Muyzer and Smalla, 1998; Diez *et al*., 2001; Fromin *et al*., 2002). Based on the relative positions and intensities of bands on DGGE patterns, Marzorati and colleagues (2008) proposed a visual analysis of the structure and diversity of microbial communities.

Here, DGGE of 16S rRNA genes was used for analyses of variation in bacterial community within 700 cm from the soil surface. Bacterial community compositions from three representative lithological zones were investigated with 16S rRNA gene clone libraries. Environmental factors, including geochemical and physical characteristics of the sediments, were determined and then related to bacterial beta diversity. The major aims of this work are to examine: (i) the vertical distribution patterns of bacteria in the shallow terrestrial subsurface with regard to alpha diversity and beta diversity, (i) the roles of contemporary environmental variables and depth in explaining vertical bacterial distribution patterns and (iii) to assess the relative importance of contemporary environmental heterogeneity and historical events on the beta diversity of bacteria.

Results

Lithological characterization

Lithologically, three main zones were identified in the core (Fig. 1A), which was consistent with the results of Zhu and colleagues (2003). Below 400 cm, the sediments were composed of caesious sludge (III: marine-originated layer; ML) deposited chiefly in a marine environment (Zhu *et al*., 2003). Slightly above that, a buff sediment layer at 380–410 cm could be distinguished. Subsamples of Hmd14 to Hmd24 were sampled from this zone. Cultural layer (II: *c*. 80–380 cm in depth; CL), as defined by Zhu and colleagues (2003), overlaid the ML. The CL consisted of grey mud in the upper portion (*c*. 80–270 cm in depth) and dark grey mud in the lower portion (*c*. 270–380 cm in depth), with the exception for a turf layer at 310–330 cm. Ten subsamples, Hmd04–Hmd13, were procured from this zone. The sediments of the surface layer (I: soil layer; SL) were dominated by peat paddy soils, from which Hmd01–Hmd03 were subsampled.

Grain size was evenly distributed throughout the depth of the core sample (8.2 μm on average) and most particles (> 92%) were less than 32 μm (Fig. 1A). Hmd07 and Hmd08 were different from the others, with approximated 45% silt and an average grain size of 4.5 μm. X-ray diffraction (XRD) results showed that all sediment samples were dominated by quartz (33–54%), muscovite (38–56%) and clinochlore (6–13%). The minor minerals were consistent with the differentiation of three main layers: a ML with calcite (2–3%), a middle layer, containing iron sulfide (FeS_2) 1–5%, and a surface layer containing neither calcite nor iron sulfide.

Sediment geochemistry

Spearman correlation analyses showed that many variables were correlated with each other (Table S1). For instance, pH was highly positively correlated with TIC ($r = 0.870$, $P < 0.001$), as well as Mg ($r = 0.900$, $P < 0.001$), Ca ($r = 0.839$, $P < 0.001$) and Mg^{2+} ($r = 0.834$, $P < 0.001$) levels. The vertical variation of most geochemical variables suggested that the SL, CL and ML zones differed from each other. For instance, pH varied from 5.50 in the surface sediment to 9.09 in the bottom sediment (Fig. 1B). In the ML, pH slightly increased, with a mean value of 8.77 ± 0.31. Above the ML, the pH was acidic, with an average value of 6.34 ± 0.57. It was noted that

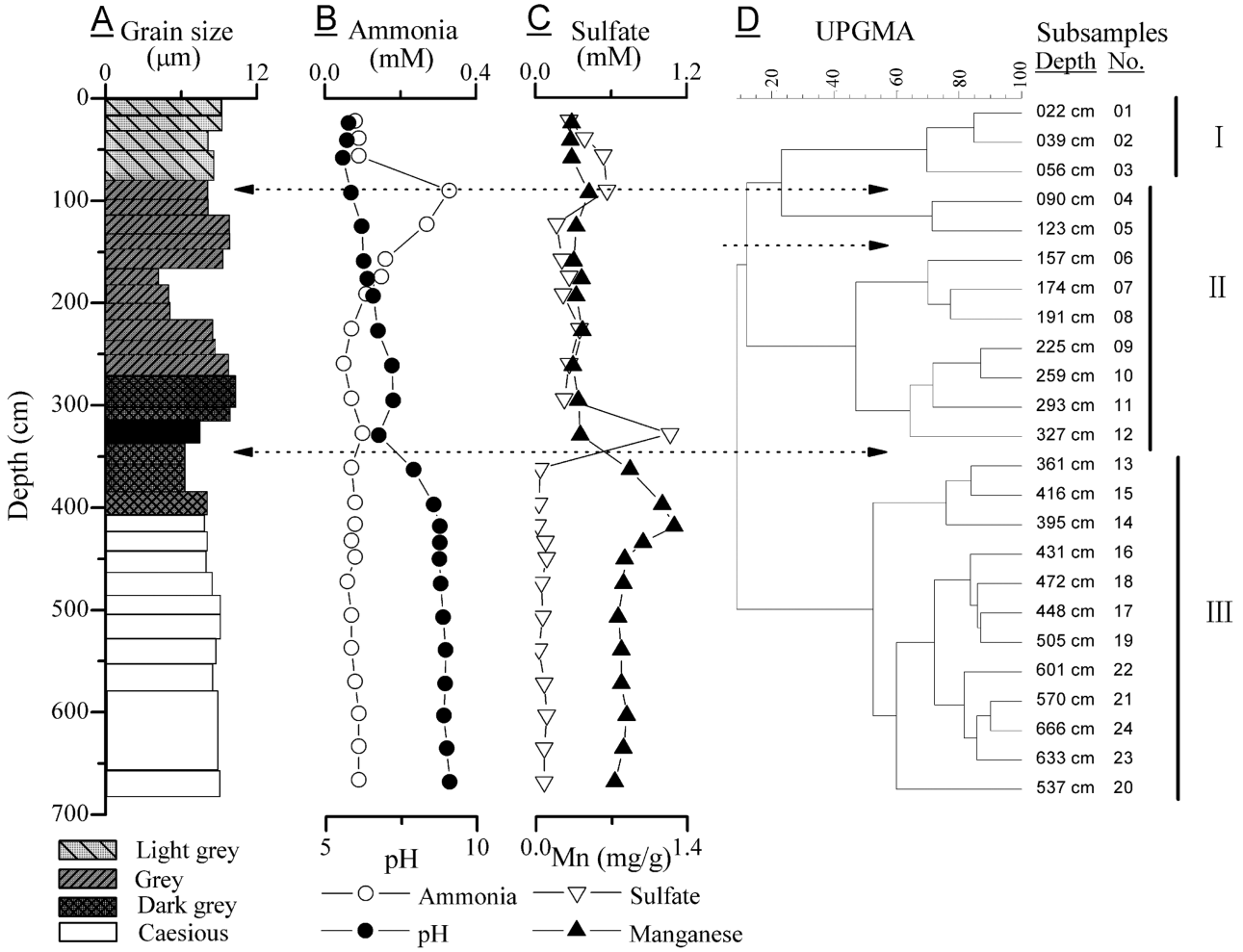

Fig. 1. A. Simplified stratigraphy. The width of bars indicates grain size.
B and C. Depth profiles of selected geochemical parameters: pH, ammonium, sulfate and Mn. Other variables are available upon request. The total subsamples were divided into three main zones, indicated by dashed lines with two arrowheads.
D. Unweighted-pair group method with arithmetic averages cluster analysis using Pearson's coefficient of DGGE band patterns, which groups the set of subsamples into four main parts, as indicated by dashed lines with right arrowheads.

ammonium was low in the SL (mean = 0.09 mM) (Fig. 1B) and the maximum ammonium concentration (0.33 mM) was found at the top of the CL; it then decreased with depth until 225 cm (Hmd09). There was only a slight variation in ammonium in the ML, with an average of 0.08 mM. Sulfate had a peak of 1.07 mM at a depth of 327 cm (Hmd12) and an average of 0.40 mM in the SL, 0.37 mM in the CL and 0.06 mM in the ML (Fig. 1C).

Dissolved organic carbon (DOC) in pore water and TOC (total organic carbon) in sediments, which are important energy resources for bacteria in the shallow terrestrial subsurface, showed variation with the depth as well (data not shown). The DOC in the water was characterized by peaks of 57.2 and 59.8 mg l^{-1} in the depths of 174 (Hmd07) and 192 cm (Hmd08), and ranged from 13.7 to 59.8 mg l^{-1}. The upper 400 cm had higher DOC levels (mean = 34.6 mg l^{-1}) than that of the ML (mean = 20.5 mg l^{-1}). The TOC in the upper 4 m of the study core was high but variable, with an average of 45.0 mg g^{-1}, and peaks of 74.6 and 74.2 mg g^{-1} at depths of 174 (Hmd07) and 327 cm (Hmd12) respectively. The opposite was true in the ML, which had a stable TOC and a lower overall concentration (mean = 7.3 mg g^{-1}).

When principal component analysis (PCA) was performed, the first two components accounted for 64.7% of the total variance of all measured environmental variables (Fig. S2). The PC1 had a strong positive relationship with pH, dissolved inorganic carbon (DIC), Ca, Mg, Mg^{2+} and dissolved inorganic phosphate (DIP), and a negative relationship with TOC, TN, sulfate and DOC. The PC2 was most strongly related to V, Cr, Be, Cu and Ni (positive relationships) and grain size, Na, Sr, Ba and TP (negative relationships) (data not shown). Overall, the PCA suggested that there were three main groups among the 24

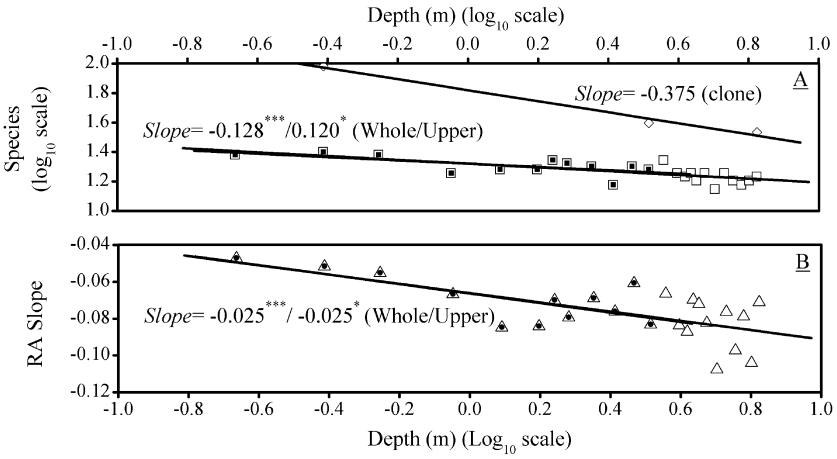

Fig. 2. Vertical alpha diversity distribution.
A. Bacterial species number profiles. Bacterial species were defined by two methods, the band numbers of DGGE fingerprints (■: upper, Hmd01–Hmd12; □: whole, Hmd01–Hmd24) and estimated number of OTUs from the three 16S rRNA clone libraries (◇: clone).
B. Rank–abundance slopes (RA slopes) profiles (●: upper, Hmd01–Hmd12; △: whole, Hmd01–Hmd24). Depth, and species or RA slopes were all $\log_{10}$-transformed. Pairwise relationship were modelled by GLM with significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$) and the slopes were shown.

subsamples: a. Hmd01–Hmd04; b. Hmd05–Hmd12, which varied more with PC2; and c. Hmd13–Hmd24, where Hmd13 was different from the others (Fig. S2).

DGGE analyses

A total of 63 bands (i.e. taxa) with different positions were detected from the DGGE fingerprints. The highest numbers of bands were detected in the three surface subsamples (24, 25 and 24 for Hmd01, Hmd02 and Hmd03 respectively) and the number of bands (14) was the lowest in subsample Hmd19. For the upper part of the core and the whole core, there were significant negative relationships between log depth and log taxa richness ($r = -0.672$, $P = 0.023$, $n = 12$, Hmd01–Hmd12; $r = -0.759$, $P < 0.0001$, $n = 24$, Hmd01–Hmd24 respectively), with slopes of -0.120 ± 0.044 and -0.128 ± 0.024 (Fig. 2A).

A rank–abundance curve was used to discern the differences between community structures among different samples based on the regression slope, with larger slope indicating greater evenness (Fig. 2B). For the upper part of the core, we found a weak but nonetheless significant linear relationship ($r = -0.694$, $P = 0.018$) between the slopes and log depth. The relationship for the whole core was also significant and inverse ($r = -0.679$, $P < 0.001$).

Unweighted-pair group method with arithmetic averages clustering indicated that there were four main community types (Fig. 1D). The first type ranged from Hmd01 to Hmd03, the second contained two subsamples (Hmd04, Hmd05), the third comprised the subsamples from Hmd06 to Hmd12 and the fourth cluster ranged from Hmd13 to Hmd24, corresponding to the ML.

Bacterial 16S rRNA gene clone libraries

Three clone libraries of 16S rRNA genes, obtained from Hmd02, Hmd12 and Hmd24, contained 134 different bacterial sequences (187 clones). A cumulative plot of all bacterial 16S rRNA gene clone libraries (Fig. 3A) was drawn from a phylogenetic analysis (data not shown). The bacterial diversity of Hmd02 was high, with sequences from 11 major phyla or class-level groups of the 52 phyla currently recognized (Rappe and Giovannoni, 2003). *Betaproteobacteria* (17.3%) and *Acidobacteria* (17.3%) were the predominant groups and about 19% of the total phyla were unclassified. In contrast to Hmd02, only six phyla were recognized in Hmd12 and Hmd24, and more than 61.0% of the total phyla belonged to *Betaproteobacteria*.

The estimated operational taxonomic units (OTUs) from the three clone libraries decreased with depth (95.9, 39.5 and 34.2 for Hmd02, Hmd12 and Hmd24 respectively) (Fig. 2A). This finding was consistent with the DGGE results. The results of LIBSHUFF analyses (Singleton *et al.*, 2001) suggested that Hmd02 was distinct from Hmd12 and Hmd24 ($P = 0.001$), while Hmd12 and Hmd24 did not differ from each other ($P = 0.083$) (data not shown).

Variation of beta diversity

Distance–decay curves for each part of the core, as well as for the whole core, were best approximated by a power–law model (Fig. 4A). A significant inverse relationship ($r = -0.518$, $n = 55$, $P < 0.001$) was observed for the lower part of the core (Hmd14–Hmd24), with a slope of -0.146 ± 0.033. For the upper part of the core and the whole core, the relationship was also significant ($r = -0.570$, $n = 66$, $P < 0.001$; $r = -0.629$, $n = 276$, $P < 0.001$ respectively), with a larger slope of -0.296 ± 0.056 and -0.338 ± 0.025 respectively. The halving distance of decay for the lower part of the core was 114.11 m, which was over one order of magnitude larger than that of the upper part of the core (10.40 m) or the whole core (7.77 m).

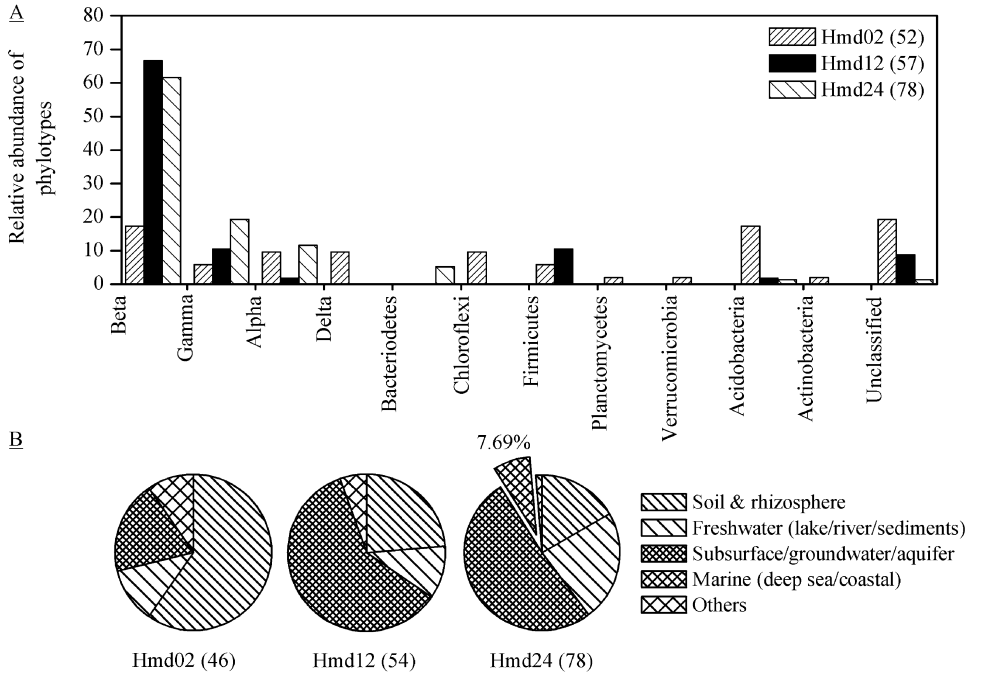

Fig. 3. Cumulative plots for 16S rRNA gene clone libraries: Hmd02, Hmd12 and Hmd24.
A. Cumulative plots of bacterial phyla. Numbers in parentheses are the numbers of sequenced clones from the three libraries. Beta, Gamma, Alpha and Delta are *Beta-*, *Gamma-*, *Alpha-* and *Deltaproteobacteria* respectively.
B. Proportion of estimated habitat affiliations of the sequences based on comparison of obtained clone sequences with GenBank Blast following the methods proposed by Souza and colleagues (2006). Each diagram indicates the habitat affiliations for each of the subsamples. Numbers in parentheses are the numbers used for affiliation estimation, and sequences with similarity < 90% were excluded from analysis.

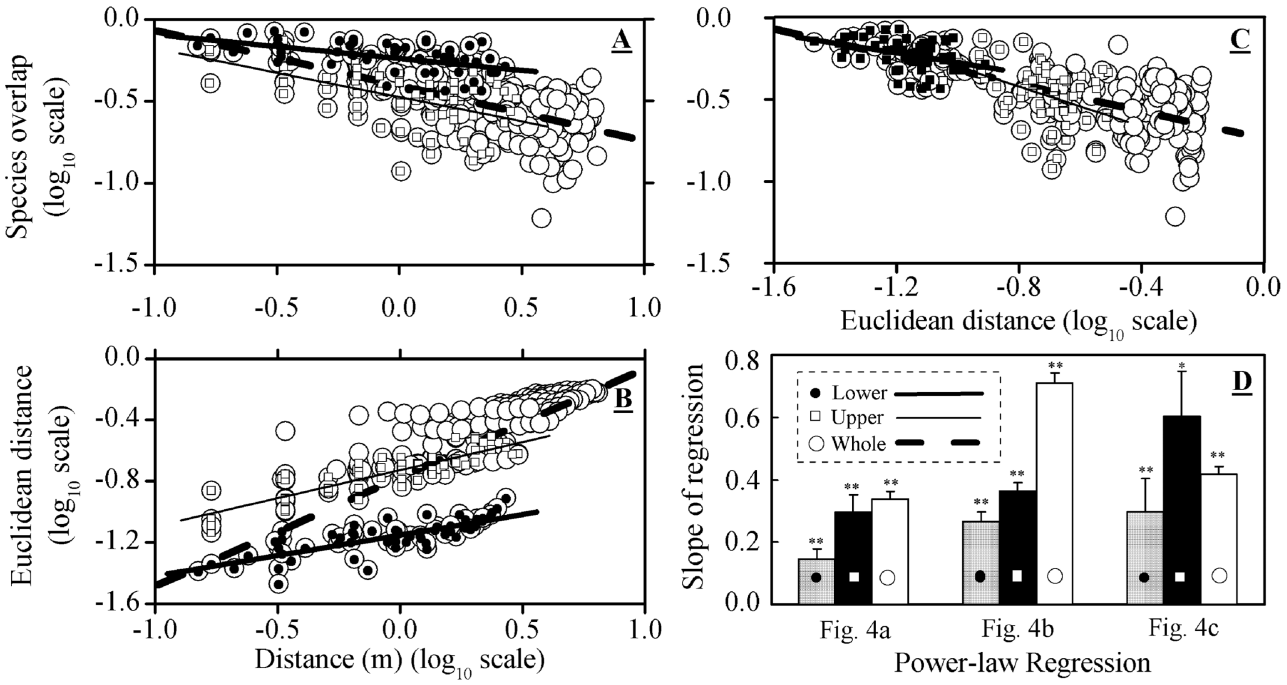

Fig. 4. Pairwise relationships between bacterial similarity, spatial distance and environmental heterogeneity (Euclidean distance) for lower (●, Hmd13–Hmd24, the lower part of the core), upper (□, Hmd01–Hmd12, the upper part of the core) and whole (○: Hmd01–Hmd24, the whole core) are indicated by solid or dotted lines in A–C: A, Sorensen similarity (species overlap) versus spatial distance; B, Euclidean distance versus spatial distance; and C, Sorensen similarity (species overlap) versus Euclidean distance. The slopes of the power–law relationship are shown in D and the significance was determined by Mantel test (*≤ 0.01, **≤ 0.001, 9999 permutations).

For the lower and upper parts of the core, as well as for the whole core, beta diversity was logarithmically related to environmental distance, and environmental and spatial distances were themselves logarithmically related (Fig. 4B and C). These relationships were all significant ($P < 0.01$) (Fig. 4D). The great variation in environment with distance for the whole core [log (Euclidean distance) = 0.709 log (spatial distance) – 0.776] revealed that the environment changed sharply from the surface to the deeper layers. The variation in environment with spatial distance in the lower part of the core was smaller than in the upper part of the core [log (Euclidean distance) = 0.267 log (spatial distance) – 1.146, log (Euclidean distance) = 0.366 log (spatial distance) – 0.729 for the lower and upper parts of the core respectively] (Fig. 4B and D). This is consistent with the PCA results (Fig. S2). Furthermore, the significant relationship between Sorensen similarity and Euclidean distance for all three analyses (Fig. 4C) revealed the controlling role of environmental distance on bacterial community patterns.

In order to tease apart the relative contributions of environmental and spatial distances in explaining the variations of beta diversity, multiple regressions were performed (Fig. 5, distance-based) (Tuomisto and Ruokolainen, 2006). Variation partitioning for the whole core showed that environmental distance alone explained 18.3%, while pure spatial distances explained less than 0.1%. For the upper part of the core, pure effects of spatial and environmental distances were only 1.9% and 3.7% respectively. However, in the lower part of the core, the pure environmental distance explained only 0.9% of the total variation, while the effect of pure spatial distance was 16.1%. This close comparison of the relative roles of environmental and spatial distances in the variation of beta diversity revealed different patterns among the three scales. For example, for the whole core, 99.9% of spatial distance variation overlapped with that of environmental distance, and environmental distance explained 72.8% of the variation of beta diversity. By contrast, in the lower part of the core, spatial distance explained most of the variation in environmental distance (94.0%).

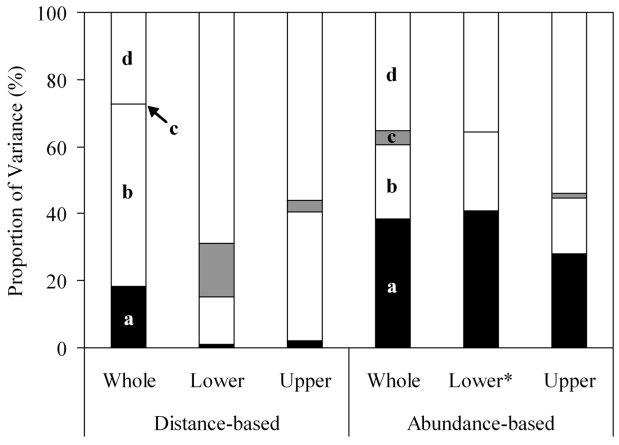

Fig. 5. Proportion of the variance in bacterial distribution explained by depth and the environmental matrix (abundance-based partition) or in variation of beta diversity explained by the spatial and environmental distances (distance-based partition).
a = environment; b = environment + spatial component; c = spatial component (*, c = zero); d = undetermined. All [a + b], [a + c], [a + b + c] were statistically significant according to the Monte Carlo permutation test ($P < 0.01$, 999 permutations) for abundance-based partitions or multiple regression ($P < 0.01$, 9999 permutations) for distance-based ones.

Factors explaining beta diversity

The explained variations in the DGGE patterns were first tested by a simple canonical correspondence analysis (CCA) model, with each factor as a single explanatory variable (Table S2). For the whole core, the geochemical variables related to mineralogy, i.e. DIC, Ca, Mg, pH, Ca^{2+}, Mg, Sr^{2+}, Na^{+}, each explained more than 17% of the total variation. When the whole core was considered, depth explained 16.6% of the total variation ($P = 0.001$). For the upper or lower parts of the core, depth was also a significant predictor of DGGE patterns.

In the unconstrained CCA, the first two axes explained 20.2% and 11.9% of the community variation in the whole core (Fig. S3A). The pH had the greatest influence on bacterial distribution. The other factors, such as Ca^{2+}, Fe, NH_4^{+}, K, Mn and DIP, also significantly structured the bacterial patterns ($P < 0.05$). For the lower and upper parts of the core, the CCA explained 46.5% and 33.5%, respectively, of the variation on the first two axes (Fig. S3B and C). The variables in explaining the most of beta diversity were Mn, Na^{+}, sulphate and pH for the axis 1, and pH, K and TOC for the axis 2.

Finally, variation partitioning methods were employed to examine the relative importance of depth and environmental factors on beta diversity (Fig. 5, abundance-based) (Tuomisto and Ruokolainen, 2006). The depth alone explained 16.6–23.5% (adjusted R^2, $P < 0.001$) of the total variations, while the shared variations accounted for much larger proportions (84–100%) of the depth fractions. In contrast, the variation in environment explained around half of the total variations (44.6–64.3%). The pure environmental effects explained remarkable proportions of the total variation (28.1–40.8%, adjusted R^2), while pure depth had only a marginal effect on beta diversity (0–4.3%).

Discussion

Distribution of alpha diversity

The positive logarithmic relationship between depth and number of taxa suggests that the upper sediments contain

more bacterial species. The notion that the deeper sediments house the fewer taxa is consistent with the species–energy theory (Wright, 1983; LaMontagne *et al*., 2003), most likely because resources are less available in lower sediments (Fierer *et al*., 2003; Lehman, 2007). Furthermore, cumulative phyla plots from the clone libraries showed that *Betaproteobacteria* are the dominant group in layers Hmd12 and Hmd24, while the surface subsample Hmd02 exhibits a community structure more evenly distributed among groups. The rank–abundance plots from DGGE confirmed this result. The finding that the bacterial communities in the lower sediments are much less diverse than in the upper layers is consistent with other studies (Zhou *et al*., 2002; Allison *et al*., 2007). These results thus imply that not only environmental heterogeneity, but also depth drive the vertical alpha diversity.

Patterns in beta diversity

For the core as a whole, the Sorensen similarity significantly declined with increasing spatial distance. This distance–decay relationship revealed that different sediment layers had unique bacterial compositions. The slope of the distance–decay relationship for over all the subsamples was −0.388, which is much larger than the reported slope for *ascomycete* fungi in soil (−0.147) (Green *et al*., 2004) or for bacteria in marsh sediments (−0.080) (Horner-Devine *et al*., 2004). To facilitate comparisons with the other studies, we also computed regression with the Jaccard-metric, and the resulting slope [i.e. −0.304 ln(Jaccard) per m of distance for the whole core] was nearly several orders of magnitudes higher than those reported for macroorganisms. For example, the rate of similarity decay was −1.85E-06 ln (Jaccard) per m of distance for tropical trees (Condit *et al*., 2002). Moreover, the mean value of halving distance for the whole core was 7.81, which is much lower than the mean value across all published studies (639.73) (Soininen *et al*., 2007) and is, to our knowledge, the lowest value reported. To conclude, our results show that the bacterial community is strongly spatially structured, indicating that the species turnover is notably high for bacteria in the shallow terrestrial subsurface. However, if this is so, what are the factors driving this beta diversity of bacteria?

The pivotal role of environmental factors in explaining beta diversity

Some recent studies have shown that pH is an important factor that influences the bacterial community in lakes (Lindstrom *et al*., 2005) and surface soils (Fierer and Jackson, 2006). However, the predominant effects of pH on microbial diversity in the subsurface have been less documented, although substantial changes of pH can also be found in subsurface conditions (Ulrich, 1998). In our study, pH increased with depth and was higher in the marine-originated sediments than in the surface layers. Based on CCA, pH was in fact the primary factor driving bacterial patterns. We would like to stress, however, that in the present study, 27 variables were significantly correlated with the bacterial DGGE patterns (Table S2). Moreover, the higher levels of DIC, Na^{+}, Mg^{2+} and Ca^{2+} in the pore water of marine sediment highlighted the importance of the dissolution of calcium carbonate in controlling the pore water chemical composition. Because pH is positively correlated with a number of other geochemical variables, we cannot totally exclude the effects of other environmental factors in controlling microbial distribution and diversity.

With regard to the terrestrial subsurface habitat, the depth is a dominant factor for determining bio-patterns in sediments, and is therefore used to describe the different subsurface habitats and microbial parameters (Lehman, 2007). Not surprisingly, our results suggest that depth can be used to predict vertical patterns of bacteria (Table S2). However, as we found a significant correlation between contemporary environmental variables and depth, it seems that depth is only a useful proxy for multiple geochemical variables at the study site. This is especially true because the partitioning variation of bacterial abundance patterns showed that the pure effect of depth is only of relatively minor importance for determining bacterial communities. Thus, our data indicate that contemporary environmental variables shape the vertical bacterial distribution in the shallow terrestrial subsurface.

Historical events in explaining variation of beta diversity

According to neutral theory (Hubbell, 2001), present-day bacterial composition in the subsurface is related to dispersal history. One piece of evidence for neutrality lies in the dominance of *Betaproteobacteria* in marine-originated sediments. Most free-living *Betaproteobacteria* were previously thought to occur exclusively in freshwater environments (Methe *et al*., 1998). Conversely, *Betaproteobacteria* are known to be far from dominant in the sediments of open marine environments (e.g. Parkes *et al*., 2005), even in the coastal sediments (e.g. Urakawa *et al*., 1999; Koizumi *et al*., 2003; Stevens *et al*., 2005; Wilms *et al*., 2006). It is therefore anticipated that there would be few *Betaproteobacteria* in the deepest marine-originated sediments. Surprisingly, we discovered the high abundance of *Betaproteobacteria* in these sediments. One possible explanation is that dispersal history or other historical events introduced *Betaproteobacteria* into the marine-originated sediments. Hmd02 and Hmd12 (17.3% and 66.7% *Betaproteobacteria* in the

clone libraries respectively) could be potential dispersal sources, through rainfall or other hydraulic connections. Therefore, once dispersal barriers were overcome, a change in community composition would be possible as a result of inoculation from the adjacent environments.

The divergence of depositional environments may also have an effect on the variation of beta diversity, because the subsurface sediment is strongly shaped by the vertical arrangement of geologic units and their weathering profiles (Lehman, 2007). Many studies on the subsurface suggest the existence of ancient genetic lineages in current bacterial communities (e.g. Fredrickson *et al*., 1995; Boivin-Jahns *et al*., 1996; Lawrence *et al*., 2000; Kovacik *et al*., 2006; Lehman, 2007). With the exception of (deep) subsurface environments, a growing body of studies has suggested that the ancient genetic lineages were present in other habitats (e.g. Souza *et al*., 2006). In our study, the most frequent environmental affiliation in the first five nearest neighbours was assigned to each sequence (Souza *et al*., 2006), and the three clone libraries showed different patterns (Fig. 3B). Our results showed that the nearest known relatives to 7.69% of clone sequences from Hmd24 were derived from marine environments, but this was not the case in the layers Hmd02 and Hmd12. This phenomenon is consistent with the fact that only the layer Hmd24 is from marine-originated sediments and suggests that the present-day bacterial communities are partly influenced by ancient depositional communities. Furthermore, the sediment core used in this study is characterized by the complex depositional environments, sea incursions and Neolithic activities (Zhu *et al*., 2003; Zong *et al*., 2007). Therefore, the high variation of depositional environments may be responsible for the high species turnover, especially in the upper part of the core.

Trade-off between environment and history

The effects of historical events (e.g. dispersal history and depositional heterogeneity, as stated above) can also be inferred from spatial patterns (Martiny *et al*., 2006). It has been shown that the vertical patterns of bacteria are highly spatially structured (Fig. 4A), which indicates a strong effect of historical events. However, the distance–decay relationship can also be affected by environment, because environmental factors are also highly spatially structured (Fig. 4C). Our environmental gradient runs from young surface soil of low pH and high carbon concentration to marine-originated sediments with high pH and low carbon concentration (Fig. 1). Interpretation of the results of variation partitioning is very complex because of intercorrelations. Also, the relative importance of spatial versus environmental distance on beta diversity seems to vary depending on which subsamples are considered. However, the probable explanation is that the variation of beta diversity is jointly controlled by environmental heterogeneity and historical events. This view is shared by many investigators studying microbes (Martiny *et al*., 2006) and macroorganisms (e.g. Condit *et al*., 2002).

We nevertheless argue that the importance of these processes does vary among the sediment layers in our study. First, in the lower part of the core, dispersal history can be considered as the main component of historical events because the ancient depositional environment was homogeneous. Dispersal limitation therefore mainly accounts for the significant effects of spatial distance, as well as for the variation of beta diversity. This is consistent with the Hubbell's neutral theory (Hubbell, 2001), which predicts that beta diversity arises from random, spatially restricted dispersal and the local dynamics of extinction and speciation. Second, for the core taken as a whole, the environmental distance is the predominant driver for the variation of beta diversity, although the spatial distance still has an effect. Great environmental differences between the upper and the lower parts of the core are consistent with the change in bacterial vertical patterns. The dispersal or other historical effects appear to be important as well, but only within the confines of vertical environmental heterogeneity. Third, in the upper part of the core, the effects of spatial and environmental distances seemed to be balanced, as contemporary environments and historical constraints are most variable at this depth. Therefore, the relative importance of spatial and environmental distances seems to depend on the heterogeneity of the system. We thereby suggest that current bacterial beta diversity is a trade-off between contemporary environment and historical events, because, for example, if the relative effect of contemporary environment is larger, then the relative effect of history must be smaller as the total variation of beta diversity is constant.

Conclusion

In the shallow terrestrial subsurface, the species richness, evenness and beta diversity of bacteria are related to depth. A closer comparison of the relative importance of depth and environment suggests, however, that beta diversity is governed mainly by environmental variables, and the pure effect of depth is relatively low. We nevertheless found that the similarity of community composition decreased rapidly with spatial distance. Notable differences in beta diversity between the upper and lower parts of the core were also observed. We argue that the high beta diversity is increased both by environmental and spatial distances, while the relative importance of environment versus history on bacterial pairwise similarity seemed to depend on heterogeneity of the system. As the

shallow terrestrial subsurface is typically characterized by high heterogeneity of environmental variables or historical events, we suggest that high beta diversity of bacteria can also be expected in other terrestrial sites. Future investigations will need to consider more sites, deeper cores and more comprehensive methods to confirm the present findings. Such investigations will certainly provide new insights into how subsurface bacterial beta diversity is shaped by contemporary environmental factors and historical events, and how bacteria are involved in subsurface biogeochemical cycling processes, as well as help in the development and testing of ecological theories.

Experimental procedures

Study areas and sampling

Our sediment samples were taken from a Neolithic site on the Ningshao plain, Hemudu town, Yuyao city, Zhejiang province (Fig. S1). The Ningshao plain was covered by seawater for thousands of years before the early Holocene (Zhao and Wu, 1986; Zong *et al*., 2007), and the thickness of sediments at this site that were deposited in a marine environment was estimated to vary from several meters to some tens of meters (Zhu *et al*., 2003). Marine-originated sediments rich in Ca (Park *et al*., 2006) were included in our cores, resulting in a long geochemical gradient.

In May of 2006, a ~700 cm intact sediment sequence was retrieved by driving stainless steel tubes (outer diameter: 11 cm, with 2.0-meter-long inner PVC tubes) into the sediment using a vibrating coring device. In the field, the sequenced cores (184, 180, 170 and 175 cm each, from the uppermost to the deepest layers) were cut longitudinally for subsampling. To recover contamination-free samples, ~0.3 cm of sediment surface, potentially containing contaminated sediments, was discarded by using a sterile knife. Next, sterile cut-off 5 ml syringes were driven into the inner portion of the core to retrieve subcores. The obtained subcores were then placed into sterile tubes for subsequent microbial analyses. The 24 subsamples retrieved displayed visible lithological texture and were designated as Hmd01, Hmd02 and so on (Fig. 1D). The shallowest subsample (Hmd01) was ~22 cm below the soil surface and the deepest, 666 cm (Fig. 1D). The corresponding 24 subsamples for geochemical analyses were also collected using the cut-off syringes. After 1 day of field sampling, all subsamples were shipped to the laboratory in a cooler with ice bags and then stored at –20°C until molecular biological and geochemical investigation. The lithological structure of the cores was recorded by visual inspection.

Sediment geochemistry

The sediment subsamples were dried in a freeze dryer. The pH was measured after shaking a sediment/water (1:10 w/v) suspension for 30 min. For the measurement of the concentration of soluble metals and nutrients concentration, 1.00 g of the freeze-dried sediment was added to 15 ml of nitrogen-bubbled deionized water in an airtight nitrogen-purged bottle. The mixture was stirred gently for 30 min, left undisturbed for 60 min and then filtered through a 0.45 μm glass filters. The filtered water was stored at –20°C for at most 2 days for further measurements. The DOC and DIC were measured by high temperature oxidation with a Shimadzu TOC analyser (model 5000). Dissolved ammonium (NH_4^+), dissolved inorganic phosphate (DIP) and dissolved nitrate (NO_3^-) were measured by a flow injection analyser (Skalar SA1000). Other dissolved anions were analysed by anion chromatography or inductively coupled plasma-atomic emission spectroscopy. The particle size distribution was measured using Malven Mastersizer 2000 (Konert and Vandenberghe, 1997). Total metals (major and trace), organic carbon, total nitrogen and total phosphorus of the sediment were analysed by using the procedures described elsewhere (Qu *et al*., 2001). Mineralogy was analysed by XRD. For the details of the measured variables, see Table S1.

DNA extraction and amplification

Genomic DNA was extracted from 0.5 g of freeze-dried sediment using the FastDNA Spin Kit for Soil (Q-BIOgen, Irvine, CA) according to the user's manual and concentrated to a volume of 100 μl. Polymerase chain reactions (PCRs, with the primers Bac27F [5′-AGA GTT TGG ATC MTG GCT CAG-3′) and Univ1492R (5′-CGG TTA CCT TGT TAC GAC TT-3′)] were performed in a 50 μl volume containing 0.2 mM dNTP, 1.5 mM $MgCl_2$, 0.2 mM of each primer, 5 μl buffer (10× Pyrobest Buffer II, Takara, China), 1.5 U Pyrobest DNA polymerase (Takara, China) and 3 μl of template DNA. The mixture was incubated in a thermal cycler (MJ Research) for 35 cycles under the following conditions: 95°C for 30 s, 58°C for 30 s and 72°C for 2 min, followed by a final elongation step at 72°C for 10 min. The PCR products were purified with a QIAprep PCR purification kit (Qiagen). For the clone library construction, dATP/dTTP were added to the end of every PCR products by an incubation at 72°C for 10 min in 53 μl volume: 48 μl purified PCR product, 1.5 mM MgCl2, 0.2 mM dNTP, 1.4 μl buffer (Fermentas) and 1 U Taq polymerase (Fermentas). For the DGGE, a 180 bp fragment of the 16S rRNA gene of bacteria was amplified on a PTC-100 thermalcycler (MJ Research, Watertown, MS) with the forward primer 338f-GC (5′-ACT CCT ACG GGA GGC AGC AG-3′) and the reverse primer 518r (5′-GTA TTA CCG CGG CTG CTG-3′) from the purified PCR products with the primers 27F and 1492R. Each 50 μl PCR mixture contained 1× PCR buffer, 200 μM nucleotide mixture, 1.5 mM $MgCl_2$, 0.5 μM (each) primer, 1 U of Taq DNA polymerase (Promega, Shanghai, China) and 1 μl template. The PCR procedure consisted of a touchdown reaction and 10 additional cycles (Muyzer *et al*., 1993). All PCR products were performed and checked by electrophoresis on a 0.8% (w/v) agarose gel.

DGGE

The DGGE was performed on a DCode Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA) following the method introduced by Muyzer and colleagues (1993). The PCR products were all separated in 8% acryla-

mide gels with a denaturing gradient from 30% to 60% (100% denaturant corresponds to 7 M urea and 40% formamide). Gels were run in 1× TAE buffer at 60°C and 75 V for 800 min, stained for 30 min in 1× TAE containing SYBR Green I (Roche, Mannheim, Germany), and documented by the Gel Doc EQ gel documentation system (Bio-Rad Laboratories, Hercules, CA). The DGGE bands were identified by the software package Bionumerics v5.0 (Applied Maths, St-Martens-Latem, Belgium).

Cloning, sequencing and phylogenetic analysis

The PCR products (2.5 μl, dATP/dTTP added) were ligated into the pGEM-T vector (Promega, Madison, WI) and transformed into TOP10 competent cells. The transformed cells were plated on Luria–Bertani plates containing 100 μl of 0.1 M IPTG, 20 μl 0.05 M Xgal and ~20 μl of 100 μg ml^{-1} ampicillin, and then incubated at 37°C for 16 h. Over 200 randomly chosen colonies per subsample were checked for inserted 16S rRNA gene sequences. Plasmid DNA with inserts was prepared with a Qiaquick PCR purification kit (QIAGEN) and BigDye 3.1 termination sequencing kit (Applied Biosystems), and sequenced with primer 27F using the ABI-3730XL DNA Analyzer.

Obtained clone sequences were manually checked for chimeras using the Ribosomal Database Project II (http://wdcm.nig.ac.jp/RDP/html/index.html), and identified chimeric sequences were removed. A total of 187 bacterial (~700 bp) sequences were subjected to phylogenetic analysis according to Jiang and colleagues (2006). Briefly, neighbour-joining phylogenies were constructed from dissimilar distances and pairwise comparisons with the Jukes–Cantor distance model using the MEGA v3.1 (molecular evolutionary genetics analysis) program. LIBSHUFF analyses were performed to compare the three clone libraries, using software available online (Singleton *et al*., 2001). The OTU assignments were performed using DOTUR (Schloss and Handelsman, 2004). Nonlinear regression, $y = a^*[1 - \exp(-b^*x)]$, was used with SigmaPlot v10.0 (Systat Software) to fit the clone distribution data and determine the maximum number of OTUs (Zhou *et al*., 2004).

Statistical analyses

Two kinds of data matrices were constructed: quantitative bacterial taxa matrices and environmental matrices (41 environmental variables, showed in Table S1). All variables, except pH, were log-transformed [most by log $(X + 1)$ and the dissolved Mn, DIP and NH_4^+ by log $(1000^*X + 1)$]. Based on lithological characterization and variation of geochemistry (see *Results*), three spatial scales were identified: all of the 24 subsamples ('Whole core'), only the upper 12 non-marine-originated subsamples (Hmd01–Hmd12, 'Upper part of the core') and the lower 11 marine-originated subsamples (Hmd14–Hmd24, 'Lower part of the core'). Hmd13 was excluded from the lower part, as it was clearly an outlier when compared with other subsamples (see PCA results, Fig. S2).

Alpha diversity along the depth. Species richness and evenness are two aspects of alpha diversity. The number of DGGE bands (intensity > 0.2%) was used to indicate the number of species or taxa (Reche *et al*., 2005). Rank–abundance plots (or dominance–diversity curves) were constructed to explore the evenness, and each plot was fitted with Generalized Linear Models (GLM) so that the relative changes in slopes with depth could be determined (Sigler *et al*., 2002; van der Gast *et al*., 2006). The estimated rank–abundance slopes and log species numbers (DGGE band numbers or estimated OTUs from 16S rRNA clone libraries) were linearly modelled with $\log_{10}$-transformed depth and the significance of correlation was calculated ($P < 0.05$).

Variation of beta diversity. We studied the variation in beta diversity using a distance-based approach (Tuomisto and Ruokolainen, 2006). The degree of shared species (qualitative Sorensen similarity) was calculated for each pair of subsamples from DGGE fingerprints (Legendre and Legendre, 1998). The distance–decay relationship (which measures how log-transformed Sorensen similarity decays with increasing log-transformed distance between pairwise subsamples) was analysed using GLM. The *P*-value for each linear regression was determined using one-tailed randomization tests on 999 permutations (null hypothesis: slope = 0). Further, halving distances (Soininen *et al*., 2007) were calculated from the logarithmic regressions, defined as the distance that halves the similarity at 1 m beneath the surface. The major advantage of the halving distance over any measure of slope is that it can be calculated for any type of relationship between similarity and distance, and the value of halving distance will not change when different units are used. This variable thus facilitates cross-study comparisons (Soininen *et al*., 2007).

In order to examine the relationship between environmental heterogeneity and spatial distance or community similarity, Euclidean distances between pairwise subsamples were calculated with the environmental matrices. A *P*-value for each relationship was computed by Mantel tests (Pearson's correlation) with the R-Package (Oksanen *et al*., 2007) using 9999 permutations.

In order to obtain the relative contributions of spatial and environmental distances on bacterial distance between pairwise sites, multiple regressions on distance matrices and community similarity matrices were run for all three scales following the procedures described by Jones and colleagues (2006), using the programs PERMUTE v3.4a5 (http://www.bio.umontreal.ca/Casgrain/en/index.html). Briefly, the total variation explained (R_T = total R^2), the total environmental variation (R_E) and the total spatial variation (R_S), estimated from multiple regressions, were used to calculate four components of the variation partitioning (Borcard *et al*., 1992): the variation of beta diversity explained by spatial distances alone = $R_T - R_E$, by spatial and environmental distances = $R_S + R_E - R_T$, by environmental distance alone = $R_T - R_S$, and a component left unexplained = $1 - R_T$.

Explaining beta diversity. We explained the degree of beta diversity using abundance-based ordination analyses (Tuomisto and Ruokolainen, 2006). Detrended correspondence analysis was used to determine that the gradient length of the species abundance matrix is larger than 2.0 along the first axis, thus implying a unimodal species–

environment relationship. We then used CCA to examine the relationship between depth or environmental heterogeneity, and bacterial distribution at all three scales. The relative contributions of variations of depth and environment in explaining beta diversity were further examined by variation partitioning based on raw data methods, which is described by several researchers (Borcard *et al*., 1992; Legendre *et al*., 2005; Peres-Neto *et al*., 2006; Ramette and Tiedje, 2007b). Because of their high correlation with pH, three variables (TIC, Ca and Mg^{2+}) were not used in variation partition analyses over the whole core. When a negative fraction (adjusted R^2) was encountered, it was interpreted as a zero (Peres-Neto *et al*., 2006). We also excluded non-significant variables using a forward selection procedure from the environmental data set to avoid overestimated variations (Okland and Eilertsen, 1994). The significance of the testable fractions was determined by 999 random permutations, using the Canoco v4.5 (Microcomputer Power, Ithaca, NY). The CCA and pCCA (partial CCA) were also performed using Canoco v4.5.

Nucleotide sequence accession numbers

The partial 16S rRNA gene sequences have been submitted to the GenBank database under Accession No. EF196931–EF197064.

Acknowledgements

We owe thanks to D.F. Zhang for help with field sampling, to X.D. Yang, X.H. Dong, P. Legendre, H. Tuomisto and A. Ramette for helpful comments on statistical analyses, to C. Tu and B. Xue for great technical assistance, to Q. Zhang, C. Zhu, J. Green, J. Martiny, C. J. van der Gast, S.M. Wang and especially two anonymous reviewers for fruitful comments on the manuscript. This work was supported by NSFC for Distinguished Young Scholars (No. 40625007) and NSFC (Nos 40572178 and 40501064).

References

Agnelli, A., Ascher, J., Corti, G., Ceccherini, M.T., Nannipieri, P., and Pietramellara, G. (2004) Distribution of microbial communities in a forest soil profile investigated by microbial biomass, soil respiration and DGGE of total and extracellular DNA. *Soil Biol Biochem* **36:** 859–868.

Allen, J.P., Atekwana, E.A., Duris, J.W., Werkema, D.D., and Rossbach, S. (2007) Microbial community structure in petroleum contaminated sediments corresponds to geophysical signatures. *Appl Environ Microbiol* **73:** 2860–2870.

Allison, V.J., Yermakov, Z., Miller, R.M., Jastrow, J.D., and Matamala, R. (2007) Using landscape and depth gradients to decouple the impact of correlated environmental variables on soil microbial community composition. *Soil Biol Biochem* **39:** 505–516.

Barns, S.M., Cain, E.C., Sommerville, L., and Kuske, C.R. (2007) Acidobacteria phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. *Appl Environ Microbiol* **73:** 3113–3116.

Boivin-Jahns, V., Ruimy, R., Bianchi, A., Daumas, S., and Christen, R. (1996) Bacterial diversity in a deep-subsurface clay environment. *Appl Environ Microbiol* **62:** 3405–3412.

Borcard, D., Legendre, P., and Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology* **73:** 1045–1055.

Condit, R., Pitman, N., Leigh, E.G., Jr, Chave, J., Terborgh, J., Foster, R.B., *et al.* (2002) Beta-diversity in tropical forest trees. *Science* **295:** 666–669.

Diez, B., Pedros-Alio, C., Marsh, T.L., and Massana, R. (2001) Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Appl Environ Microbio* **67:** 2942–2951.

Fierer, N., and Jackson, R.B. (2006) From the cover: the diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103:** 626–631.

Fierer, N., Schimel, J.P., and Holden, P.A. (2003) Variations in microbial community composition through two soil depth profiles. *Soil Biol Biochem* **35:** 167–176.

Franklin, R., and Mills, A. (2007) Introduction. In *The Spatial Distribution of Microbes in the Environment*. Franklin, R., and Mills, A. (eds). Dordrecht, the Netherlands: Springer, pp. 1–30.

Fredrickson, J.K., McKinley, J.P., Nierzwicki-Bauer, S.A., White, D.C., Ringelberg, D.B., Rawson, S.A., *et al.* (1995) Microbial community structure and biogeochemistry of Miocene subsurface sediments: Implications for long-term microbial survival. *Mol Ecol* **4:** 619–626.

Fromin, N., Hamelin, J., Tarnawski, S., Roesti, D., Jourdain-Miserez, K., Forestier, N., *et al.* (2002) Statistical analysis of denaturing gel electrophoresis (DGE) fingerprinting patterns. *Environ Microbiol* **4:** 634–643.

Fry, J.C., Webster, G., Cragg, B.A., Weightman, A.J., and Parkes, R.J. (2006) Analysis of DGGE profiles to explore the relationship between prokaryotic community composition and biogeochemical processes in deep subseafloor sediments from the Peru Margin. *FEMS Microbiol Ecol* **58:** 86–98.

van der Gast, C.J., Jefferson, B., Reid, E., Robinson, T., Bailey, M.J., Judd, S.J., and Thompson, I.P. (2006) Bacterial diversity is determined by volume in membrane bioreactors. *Environ Microbiol* **8:** 1048–1055.

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

Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747–750.

van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., *et al.* (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proc Natl Acad Sci USA* **104:** 20404–20409.

Hansel, C.M., Fendorf, S., Jardine, P.M., and Francis, C.A. (2008) Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol* **74:** 1620–1633.

Holden, P.A., and Fierer, N. (2005) Microbial processes in the vadose zone. *Vadose Zone J* **4:** 1–21.

Horner-Devine, M.C., Lage, M., Hughes, J.B., and Bohannan, B.J.M. (2004) A taxa-area relationship for bacteria. *Nature* **432:** 750–753.

Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ, USA: Princeton University Press.

Jiang, H., Dong, H., and Zhang, G., Yu, B., Chapman, L.R., and Fields, M.W. (2006) Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. *Appl Environ Microbiol* **72:** 3832–3845.

Jones, M.M., Tuomisto, H., Clark, D.B., and Olivas, P. (2006) Effects of mesoscale environmental heterogeneity and dispersal limitation on floristic variation in rain forest ferns. *J Ecol* **94:** 181–195.

Koizumi, Y., Kojima, H., and Fukui, M. (2003) Characterization of depth-related microbial community structure in lake sediment by denaturing gradient gel electrophoresis of amplified 16S rDNA and reversely transcribed 16S rRNA fragments. *FEMS Microbiol Ecol* **46:** 147–157.

Konert, M., and Vandenberghe, J.E.F. (1997) Comparison of laser grain size analysis with pipette and sieve analysis: a solution for the underestimation of the clay fraction. *Sedimentology* **44:** 523–535.

Kovacik, W.P., Takai, K., Mormile, M.R., McKinley, J.P., Brockman, F.J., Fredrickson, J.K., and Holben, W.E. (2006) Molecular analysis of deep subsurface Cretaceous rock indicates abundant Fe (III)- and S^0-reducing bacteria in a sulfate-rich environment. *Environ Microbiol* **8:** 141–155.

LaMontagne, M.G., Schimel, J.P., and Holden, P.A. (2003) Comparison of subsurface and surface soil bacterial communities in california grassland as assessed by terminal restriction fragment length polymorphisms of PCR-amplified 16S rRNA genes. *Microb Ecol* **46:** 216–217.

Lawrence, J., Hendry, M., Wassenaar, L., Germida, J., Wolfaardt, G., Fortin, N., and Greer, C. (2000) Distribution and biogeochemical importance of bacterial populations in a thick clay-rich aquitard system. *Microb Ecol* **40:** 273–291.

Legendre, P., and Legendre, L. (1998) *Numerical Ecology*. New York, NY, USA: Elsevier.

Legendre, P., Borcard, D., and Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecol Monogr* **75:** 435–450.

Lehman, R. (2007) Microbial distributions and their potential controlling factors in terrestrial subsurface environments. In *The Spatial Distribution of Microbes in the Environment*. Franklin, R. and Mills, A. (eds). Dordrecht, the Netherlands: Springer, pp. 135–178.

Lindstrom, E.S., Kamst-Van Agterveld, M.P., and Zwart, G. (2005) Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl Environ Microbiol* **71:** 8201–8206.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007) Quantitative and qualitative β 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.

Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D., and Verstraete, W. (2008) How to get more out of molecular fingerprints: practical tools for microbial ecology. *Environ Microbiol* **10:** 1571–1581.

Methe, B.A., Hiorns, W.D., and Zehr, J.P. (1998) Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George and six other adirondack lakes. *Limnol Oceanogr* **43:** 368–374.

Muyzer, G., and Smalla, K. (1998) Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek* **73:** 127–141.

Muyzer, G., de Waal, E.C., and Uitterlinden, A.G. (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59:** 695–700.

Okland, R.H., and Eilertsen, O. (1994) Canonical correspondence analysis with variation partitioning: some comments and an application. *J Veg Sci* **5:** 117–126.

Oksanen, J., Kindt, R., Legendre, P., and O'Hara, R.B. (2007) *Vegan: community ecology package*, *R. Package Version 1.8–6.* [WWW document]. URL http://cran.r-project.org/.

Park, J., Sanford, R.A., and Bethke, C.M. (2006) Geochemical and microbiological zonation of the Middendorf aquifer, South Carolina. *Chem Geol* **230:** 88–104.

Parkes, R.J., Webster, G., Cragg, B.A., Weightman, A.J., Newberry, C.J., Ferdelman, T.G., *et al.* (2005) Deep subseafloor prokaryotes stimulated at interfaces over geological time. *Nature* **436:** 390–394.

Peres-Neto, P.R., Legendre, P., Dray, S., and Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87:** 2614–2625.

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

Qu, W.C., Mike, D., and Wang, S.M. (2001) Multivariate analysis of heavy metal and nutrient concentrations in sediments of Taihu Lake, China. *Hydrobiologia* **450:** 83–89.

Ramette, A. (2007) Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* **62:** 142–160.

Ramette, A., and Tiedje, J. (2007a) Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microbial Ecol* **53:** 197–207.

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

Rappe, M.S., and Giovannoni, S.J. (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57:** 369–394.

Reche, I., Pulido-Villena, E., Morales-Baquero, R., and Casamayor, E.O. (2005) Does ecosystem size determine aquatic bacterial richness? *Ecology* **86:** 1715–1722.

Schloss, P.D., and Handelsman, J. (2004) Status of the microbial census. *Microbiol Mol Biol Rev* **68:** 686–691.

Shade, A., Jones, S.E., and McMahon, K.D. (2008) The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environ Microbiol* **10:** 1057–1067.

Sigler, W.V., Crivii, S., and Zeyer, J. (2002) Bacterial succession in glacial forefield soils characterized by community

structure, activity and opportunistic growth dynamics. *Microb Ecol* **44:** 306–316.
Singleton, D.R., Furlong, M.A., Rathbun, S.L., and Whitman, W.B. (2001) Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* **67:** 4374–4376.
Soininen, J., McDonald, R., and Hillebrand, H. (2007) The distance decay of similarity in ecological communities. *Ecography* **30:** 3–12.
Souza, V., Espinosa-Asuar, L., Escalante, A.E., Eguiarte, L.E., Farmer, J., Forney, L., *et al.* (2006) An endangered oasis of aquatic microbial biodiversity in the Chihuahuan desert. *Proc Natl Acad Sci USA* **103:** 6565–6570.
Stevens, H., Stubner, M., Simon, M., and Brinkhoff, T. (2005) Phylogeny of *Proteobacteria* and *Bacteroidetes* from oxic habitats of a tidal flat ecosystem. *FEMS Microbiol Ecol* **54:** 351–365.
Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O., and Reysenbach, A.-L. (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles. *Environ Microbiol* **10:** 1681–1689.
Tuomisto, H., and Ruokolainen, K. (2006) Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology* **87:** 2697–2708.
Ulrich, G.A. (1998) Sulfur cycling in the terrestrial subsurface: commensal interactions, spatial scales, and microbial heterogeneity. *Microb Ecol* **36:** 141–151.
Urakawa, H., Kita-Tsukamoto, K., and Ohwada, K. (1999) Microbial diversity in marine sediments from Sagami Bay and Tokyo Bay, Japan, as determined by 16S rRNA gene analysis. *Microbiology* **145:** 3305–3315.
Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci* **95:** 6578–6583.
Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon* **21:** 213–251.
Wilms, R., Koepke, B., Sass, H., Chang, T.S., Cypionka, H., and Engelen, B. (2006) Deep biosphere-related bacteria within the subsurface of tidal flat sediments. *Environ Microbiol* **8:** 709–719.
Woodcock, S., Curtis, T.P., Head, I.M., Lunn, M., and Sloan, W.T. (2006) Taxa–area relationships for microbes: the unsampled and the unseen. *Ecol Lett* **9:** 805–812.
Wright, D.H. (1983) Species-energy theory: an extension of species-area theory. *Oikos* **41:** 496–506.
Zhao, S., and Wu, W.T. (1986) Early neolithic Hemudu Culture along the Hangzhou Estuary and the origin of domestic paddy rice in China. *Asian Perspect (Honolulu)* **27:** 29–34.
Zhou, J., Xia, B., Treves, D.S., Wu, L.Y., Marsh, T.L., O'Neill, R.V., *et al.* (2002) Spatial and resource factors influencing high microbial diversity in soil. *Appl Environ Microbiol* **68:** 326–334.
Zhou, J., Xia, B., Huang, H., Palumbo, A.V., and Tiedje, J.M. (2004) Microbial diversity and heterogeneity in sandy subsurface soils. *Appl Environ Microbiol* **70:** 1723–1734.
Zhu, C., Zheng, C.G., Ma, C.M., Yang, X.X., Gao, X.Z., Wang, H.M., and Shao, J.H. (2003) On the Holocene sea-level highstand along the Yangtze Delta and Ningshao Plain, East China. *Chin Sci Bull* **48:** 2672–2683.
Zong, Y., Chen, Z., Innes, J.B., Chen, C., Wang, Z., and Wang, H. (2007) Fire and flood management of coastal swamp enabled first rice paddy cultivation in east China. *Nature* **449:** 459–462.

Supplementary material

The following supplementary material is available for this article online:

Table S1. Spearman correlation matrix of geochemical or spatial variables.

Table S2. Variation explained by single variable: each factor was imported into the CCA model as a single explanatory variable, and the significance was determined by 999 unrestricted Monte Carlo permutations (*$\leq$ 0.05, **$\leq$ 0.01, ***$\leq$ 0.001). The abundance matrices were used.

Fig. S1. Sampling site. The sediment sequence was collected from the archaeological site of Hemudu, Yuyao City, Zhejiang Province, China. The drilling site is 1 km north of the Hemudu Site Museum and it is shown on the map with a black triangle.

Fig. S2. PCA plots of the first two principal components of all environmental data. Values on the axes indicate the percentages of total variation explained by each axis. The three clusters are shown by dotted lines and the numbers 1–24 represent corresponding subsamples Hmd01–Hmd24. Upper: Hmd01–Hme12, the upper part of the core; Lower: Hmd14–Hmd24, the lower part of the core.

Fig. S3. CCA biplots for the bacterial abundance matrix from the DGGE profile. A, the whole core, Hmd01–Hmd24; B, the upper part of the core, Hmd01–Hmd12; C, the lower part of the core, Hmd14–Hmd24. All auto-selected environmental variables are statistically significant in contributing to the CCA model according to the Monte Carlo permutation test ($P < 0.05$, 999 permutations). The labelled numbers (1–24) stand for subsamples Hmd01–Hmd24.

This material is available as part of the online article from http://www.blackwell-synergy.com

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

