ORIGINAL ARTICLE

Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes

Jianjun Wang[1,2], Ji Shen[1], Yucheng Wu[3], Chen Tu[4], Janne Soininen[5], James C Stegen[6], Jizheng He[2], Xingqi Liu[1], Lu Zhang[1] and Enlou Zhang[1]

[1]*State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China;* [2]*State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China;* [3]*State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China;* [4]*Key Laboratory of Coastal Zone Environmental Processes, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, China;* [5]*Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland and* [6]*Fundamental and Computational Sciences Directorate, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA*

Increasing evidence has emerged for non-random spatial distributions of microbes, but knowledge of the processes that cause variation in microbial assemblage among ecosystems is lacking. For instance, some studies showed that deterministic processes such as habitat specialization are important, while other studies hold that bacterial communities are assembled by stochastic forces. Here we examine the relative influence of deterministic and stochastic processes for bacterial communities from subsurface environments, stream biofilm, lake water, lake sediment and soil using pyrosequencing of the 16S ribosomal RNA gene. We show that there is a general pattern in phylogenetic signal in species ecological niches across recent evolutionary time for all studied habitats, enabling us to infer the influences of community assembly processes from patterns of phylogenetic turnover in community composition. The phylogenetic dissimilarities among-habitat types were significantly higher than within them, and the communities were clustered according to their original habitat types. For communities within-habitat types, the highest phylogenetic turnover rate through space was observed in subsurface environments, followed by stream biofilm on mountainsides, whereas the sediment assemblages across regional scales showed the lowest turnover rate. Quantifying phylogenetic turnover as the deviation from a null expectation suggested that *measured* environmental variables imposed strong selection on bacterial communities for nearly all sample groups. For three sample groups, spatial distance reflected *unmeasured* environmental variables that impose selection, as opposed to spatial isolation. Such characterization of spatial and environmental variables proved essential for proper interpretation of partial Mantel results based on observed beta diversity metrics. In summary, our results clearly indicate a dominant role of deterministic processes on bacterial assemblages and highlight that bacteria show strong habitat associations that have likely emerged through evolutionary adaptation.

Introduction

A major theme of ecological research is characterizing the processes underlying spatial variation in biotic communities (that is, beta diversity) across

Earth's ecosystems (Hubbell, 2001; Martiny *et al.*, 2006; Anderson *et al.*, 2011). It has been recognized that habitat specialization resulting from the adaptive evolution by means of natural selection (Darwin, 1859) has a pivotal role in determining community composition (for example, Graham and Fine, 2008; Cavender-Bares *et al.*, 2009). This is a classic deterministic process driven by contemporary environmental heterogeneity. However, variation in communities is also influenced by stochastic processes such as dispersal limitation, mass effects and random demographics (Hubbell, 2001; Leibold

et al., 2004; Cottenie, 2005; Martiny *et al.*, 2006; Vellend, 2010; Chase and Myers, 2011).

Although stochastic and deterministic processes have long been studied in plant and animal systems, extensive study of these processes in microbial communities has emerged only in the past decade (Besemer *et al.*, 2012; Hanson *et al.*, 2012 and references therein). These studies provide strong evidence of biogeographical patterns for microbes, the distance–decay relationship being one example (for example, Martiny *et al.*, 2006). However, still little is known about the processes that underlie non-random distributions of microbes.

Meta-analysis have shown that bacterial communities differ substantially among-habitat types (for example, Lozupone and Knight, 2007; Delmont *et al.*, 2011), with salinity being a key factor structuring microbial communities (Lozupone and Knight, 2007). These patterns suggest a dominant role of habitat specialization for microbial community assembly. However, the inconsistency of public microbial gene sequences or experimental methods and inaccessibility of consistent environmental information prevent more detailed profiles on the relative influence of deterministic and stochastic processes. Further, processes governing between-habitat variation in community composition may differ from those responsible for variation within habitats. Until now, few studies on microbial communities have characterized biogeographical patterns and underlying processes across- and within-habitat types.

Even with consistent microbial and environmental data sets, there are important factors to be considered when charactering underlying processes. For example, contemporary environmental variables measured in the field are typically spatially autocorrelated and some ecologically important variables may be left unmeasured, both factors complicate the inferences related to the relative influences of stochastic and deterministic processes. However, methods from macro-organism ecology may provide good solutions to these challenges. In particular, coupling the spatial variation of phylogenetic community structure (Webb *et al.*, 2002) with null models can help characterize the relative influences of deterministic and stochastic processes (Graham and Fine, 2008; Chase and Myers, 2011; Stegen *et al.*, 2012).

Inferring underlying ecological processes using phylogenetic information requires phylogenetic signal in habitat association (Losos, 2008; Cavender-Bares *et al.*, 2009). Following detection of phylogenetic signal, within a phylogenetic framework (Graham and Fine, 2008; Chase and Myers, 2011; Fine and Kembel, 2011; Stegen *et al.*, 2012), we propose a three-step procedure to characterize the relative influences of deterministic and stochastic processes.

First, we test for an influence of deterministic processes by comparing observed phylogenetic turnover between assemblages to a stochastic expectation that controls for observed turnover in taxonomic composition. Significant deviations from the stochastic expectation indicate that deterministic processes such as environmental filtering strongly influence community composition.

Second, we use a new approach to reveal influences of *unmeasured* environmental variables by relating phylogenetic null model deviations to environmental and spatial distances. A significant relationship with spatial distances, after controlling for environmental distances, suggests that *unmeasured* environmental variables impose deterministic processes that overwhelm any influences of stochastic processes. This is because stochastic processes should not cause phylogenetic null model deviations (Hardy, 2008).

Third, we evaluate the relative influences of deterministic and stochastic processes by considering results from the previous step and additional analyses that relate observed phylogenetic turnover to spatial and environmental distances: a stronger influence of stochastic versus deterministic processes can be inferred if null model deviations are not related to spatial distance and if observed phylogenetic turnover increases with spatial distance to a greater degree than environmental distance. We infer that deterministic processes have more influence than stochastic processes if observed phylogenetic turnover correlates more strongly with environmental than with spatial distances (Stegen and Hurlbert, 2011). We also infer a greater influence of deterministic processes if null model deviations increase with spatial distance, which suggests that any influence of spatial isolation is overwhelmed by environmental selection imposed by *unmeasured* environmental variables.

Here, we analyzed bacterial assemblages from aquatic and terrestrial subsurface environments, as well as lake water, stream biofilm, lake sediment and soil using pyrosequencing of the 16S ribosomal RNA gene. It should be noted that although previous microbial work has examined most of Earth's major ecosystems, aquatic and terrestrial subsurface environments have received little attention. Compared with previous meta-analyses that incorporate among-habitat comparisons (for example, Lozupone and Knight, 2007), we examine a broad range of habitats using more consistent methods to characterize community and environmental data.

In addition to evaluating phylogenetic signal and inferring ecological processes across habitats, our study aims to (1) examine whether bacteria show clear habitat associations, that is, whether community composition differs among sampled habitat types; and (2) compare among-habitat types the rate at which community composition turns over through space.

Materials and methods

Data collection
We collected bacterial samples from aquatic or terrestrial subsurface sediment, lake water, stream

Table 1 Brief description of the sample groups with different habitat types

Sample group	Location	Sample number[a]	Sampling date	Habitat types	Description
KL1	31.441°N, 120.814°E	12	2010-01	Surface sediments (0–1 cm)	Kuilai Lake
KL7	31.441°N, 120.814°E	12	2010-07	Surface sediments (0–1 cm)	Kuilai Lake
ThS	31.192°N, 120.106°E	27	2010-06	Surface sediments (0–1 cm)	Taihu Lake
SC	26°–34°N, 98°–103°E	25	2009-10	Surface sediments (0–1 cm)	Lakes in Sichuang Regions
KSS	35.737°N, 92.881°E	2	2010-09	Surface sediments (0–1 cm)	Kusai Lake[b]
KS	35.737°N, 92.881°E	28	2010-09	Subsurface sediments[c]	Kusai Lake
LG	27.713°N, 100.772°E	20	2008-10	Subsurface sediments	Lugu Lake
NJ	32.061°N, 118.789°E	5	2010-11	Subsurface soil[c]	A park in Nanjing City
ZJ	29.944°N, 121.015°E	7	2010-06	Surface soils (0–4 cm)	Fields in Zhejiang Province
HX	35.782°N, 92.974°E	10	2010-09	Surface soils (0–4 cm)	Qinghai–Tibetan plateau
STR	26.843°N, 99.964°E	24	2009-11	Stream biofilm	Laojun Mountains
ThW	31.192°N, 120.106°E	27	2010-06	Lake water	Taihu Lake

[a]Non-metric multidimensional scaling was applied for all sample groups. Mantel correlograms (that is, the test of phylogenetic signal) and variations in phylobetadiversity (Unifrac, mean nearest taxon distance separating OTUs in two communities (MNTD) or standardized effect size of MNTD) were only examined for sample groups with ⩾ 10 samples.
[b]Kusai Lake is a saline lake with a salinity of ∼17‰. Owing to the paleo-environmental changes in the Kusai Lake regions, the lake salinity varied in the past 2000 years (Liu *et al.*, 2009). The other lakes or streams (Wang *et al.*, 2011) are freshwater.
[c]Subsurface sediments: 1 cm below lake floor; subsurface soils: 10 cm below land surface.

biofilm, lake surface sediment and soil (Table 1). Sequences generated from pyrosequencing of bacterial 16S ribosomal RNA gene amplicons were processed using QIIME pipeline (v1.2) (Caporaso *et al.*, 2010). Details related to sample collection, DNA sequencing and analyses are described in the supplementary text and Wang *et al.* (2012a). Operational taxonomic units (OTUs) were defined using a 97% sequence similarity cutoff. We accounted for differences in sampling effort among samples by randomly subsampling 1000 sequences per sample for 1000 times.

Statistical analyses
To evaluate phylogenetic signal across a range of phylogenetic depths, we used Mantel correlograms with 999 randomizations for significance tests (Oden and Sokal, 1986; Diniz-Filho *et al.*, 2010) with the function 'mantel.correlog' in the R package Vegan v2.0-2 (http://vegan.r-forge.r-project.org). We partitioned phylogenetic distances into classes (that is, evolutionary time steps; here 0.02 units) and within each distance class we found the correlation coefficient relating between-OTU phylogenetic distances to environmental-optimum distances (Diniz-Filho *et al.*, 2010). This method has the advantage of characterizing shifts in phylogenetic signal across phylogenetic distance classes (Diniz-Filho *et al.*, 2010). An environmental-optimum for each OTU was found for each environmental variable as in Stegen *et al.* (2012). Between-OTU environmental-optimum differences were calculated as Euclidean distances using optima for all the environmental variables.

To quantify phylogenetic turnover in community composition between a given pair of samples (that is 'phylogenetic beta diversity' or 'phylobetadiversity'), we used unweighted Unifrac (Lozupone and Knight, 2005) and the mean nearest taxon distance

separating OTUs in two communities (betaMNTD) (Fine and Kembel, 2011; Stegen *et al.*, 2012). BetaMNTD is the mean phylogenetic distance to the closest relative in a paired community for all taxa (Fine and Kembel, 2011) and is sensitive to the changes of lineages close to the phylogenetic tips.

We performed non-metric multidimensional scaling based on unweighted Unifrac and betaMNTD to depict community composition in two dimensions. To test the hypothesis that habitat types structure the distribution of bacteria, permutational multivariate analysis of variance was used (Anderson, 2001).

For each habitat, a standardized effect size (ses.betaMNTD) was computed as the number of standard deviations that observed betaMNTD departed from the mean of null distribution (999 null iterations) based on random shuffling of OTU labels across the tips of the phylogeny (Hardy, 2008; Fine and Kembel, 2011; Stegen *et al.*, 2012). This randomization holds constant observed species richness, species occupancy and species turnover. It therefore provides an expected level of betaMNTD given observed species richness, occupancy and turnover. The absolute magnitude of ses.betaMNTD reflects the influence of deterministic processes; the larger the magnitude, the greater the influence of deterministic, niche-based processes.

To examine variation in phylobetadiversity within the nine sample groups (those with sample number ⩾ 10, Table 1), we used a distance-based approach (Martiny *et al.*, 2006; Tuomisto and Ruokolainen, 2006) akin to distance–decay analysis where phylogenetic dissimilarity is related to spatial and environmental distance among sampled communities. Environmental distance was measured as Euclidean distance using all environmental variables standardized to have a mean of zero and a standard deviation of one. Phylobetadiversity was regressed against spatial or environmental distances

using a Gaussian generalized linear model. Significance was determined using Mantel tests (Spearman's correlation) with 9999 permutations (Legendre and Legendre, 1998). We used an analysis of covariance with permutation to test the hypothesis that the regression slopes do not differ among the sample groups. Further, partial Mantel tests were used to assess the relationship between phylogenetic turnover and spatial or measured environmental distance after accounting for measured environmental distance or spatial distance, and the significance was assessed using 9999 permutations. These analyses were performed in the R environment with Picante v1.4 (http://picante.r-forge. r-project.org) and Vegan v2.0-2

Finally, we inferred the underlying processes following the three steps listed in the Introduction section. For instance, the variation in the magnitude of ses.betaMNTD should be driven primarily by variation in the influence of deterministic processes as deviations from the betaMNTD null model expectation are primarily due to niche-based processes (Hardy, 2008). Significant increases in ses.betaMNTD with increasing environmental distances therefore implies that the influence of niche-based processes grows with increasingly large shifts in *measured* environmental conditions (Fine

and Kembel, 2011; Stegen *et al.*, 2012). Importantly, significant increases in ses.betaMNTD with increasing spatial distances implies that the influence of niche-based processes grows with increasingly large shifts in spatially structured, but *unmeasured* environmental conditions.

Results

Mantel correlograms consistently showed significant positive correlations across short phylogenetic distances for all nine sample groups ($P < 0.05$, Figures 1a–i). The phylogenetic distance across which there was significant phylogenetic signal varied from 10% to 30% of the maximum phylogenetic distance within each phylogeny. For nearly all sample groups (except for KL1 in Figure 1a), there were significant negative correlations at intermediate phylogenetic distances ($P < 0.05$) and non-significant relationships across longer distances (Figures 1b–i).

Non-metric multidimensional scaling using both unweighted Unifrac and betaMNTD showed that samples were phylogenetically segregated by habitat type (Figures 2a and c). For each sample group, the Unifrac metric showed higher values than
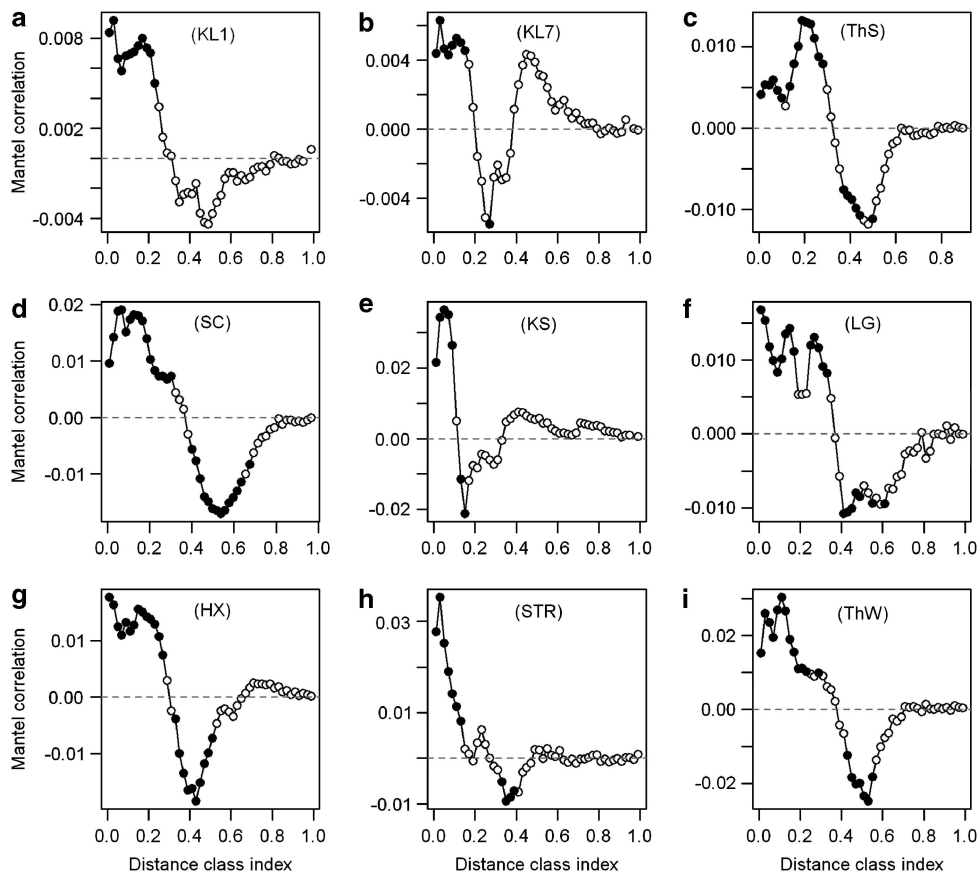

Figure 1 (a-i) Pearson correlation resulting from Mantel correlogram between the pairwise matrix of OTU niche distances and phylogenetic distances (with Jukes–Cantor model) for each sample group with 9999 permutations. Significant correlations ($P < 0.05$, solid circles) indicate phylogenetic signal in species ecological niches, and were consistently found across short phylogenetic distances for all sample groups.
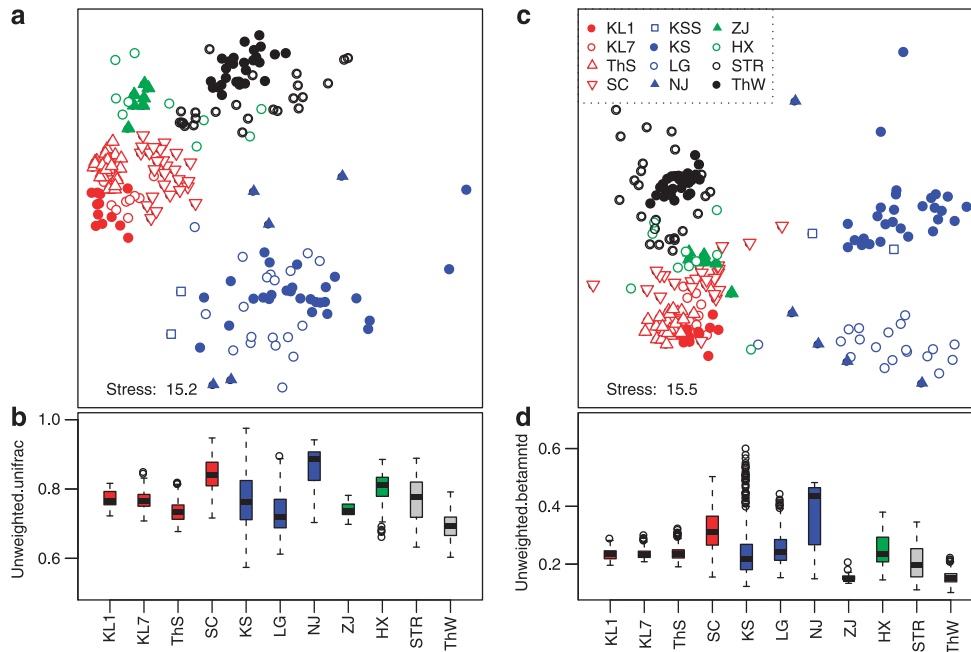
Figure 2 Non-metric multidimensional scaling plots (**a**, **c**), or boxplots (**b**, **d**) of community dissimilarities within-habitat groups using unweighted Unifrac and betaMNTD, respectively. The samples are colored by habitat groups. The habitat types include lake surface sediments (KL, ThS and SC), lake subsurface sediments/soils (KS LG and NJ), surface soils (ZJ and HX), stream biofilm (STR) and lake water (ThW). Two surface sediments from Kuisai Lake (KSS) are not included in KS group. KL1 and KL7 indicate the surface sediments sampled from Kuilei Lake in January and July, respectively. More details on the abbreviation of groups, see Table 1 and the supplementary text. Components of the box are: top of the box, upper hinge; midline of box, median; bottom of box, lower hinge; bars, 1.5 times length of box (1.5 times the horizontal spread); dots, values that are > or <1.5 times the horizontal spread of the distribution, plus the upper or lower hinge.

betaMNTD and the range of unweighted Unifrac was smaller than that of betaMNTD (Figures 2b and d). Community differences among-habitat types were also observed across short spatial distances, that is, there were large differences in community composition between surface sediments and lake water in Taihu Lake (Supplementary Figure S1A) or between soils and lake sediments in Kusai Lake regions (Supplementary Figure S1B). Analyses of permutational multivariate analysis of variance showed that habitat type explained 50.0% and 21.2% of the variation in community composition, using betaMNTD and Unifrac metrics, respectively ($P < 0.001$, 9999 permutations).

A plot of pairwise phylobetadiversity versus spatial distance showed that there was a significant distance–decay relationship for most of the sample groups: six out of nine using Unifrac (Figure 3, Supplementary Table S1) or five out of nine using betaMNTD (Supplementary Figure S2, Supplementary Table S2). For both metrics, the slope of this relationship varied significantly among most groups ($P < 0.01$), with the following decreasing order in spatial turnover rates: KS > LG > STR > ThS > ThW > SC.

Except for sample groups KL1, KL7 and ZJ, mean values of ses.betaMNTD were significantly different from the expected value of zero ($P < 0.001$, *t*-test; Supplementary Figure S3). After controlling for spatial distance, environmental distance was

significantly correlated with ses.betaMNTD within seven sample groups (Table 2). Spatial distance was significantly ($P < 0.05$) correlated with ses.beta-MNTD for five sample groups (Supplementary Figure S4). However, after controlling for environmental distance, spatial distance was significantly correlated with ses.betaMNTD only in ThS, KS and STR (partial Mantel test, $P < 0.05$; Table 2).

There were six sample groups in which ses.betaMNTD was not significantly related to spatial distance after controlling for environmental distance (Table 2). In five out of these six groups unweighted Unifrac and betaMNTD were more strongly related to environmental distance (after controlling for spatial distance) than to spatial distance (after controlling for environmental distance; Supplementary Table S1, S2). The exception was group HX, for which all beta diversity metrics showed no relationship to spatial or environmental distances.

Discussion

Here we studied a broad range of ecosystems to assess patterns of microbial community composition and the processes that underlie these patterns. To do so, we have characterized patterns of spatial turnover in the phylogenetic composition of microbial communities, and have inferred processes by comparing phylogenetic turnover to null model
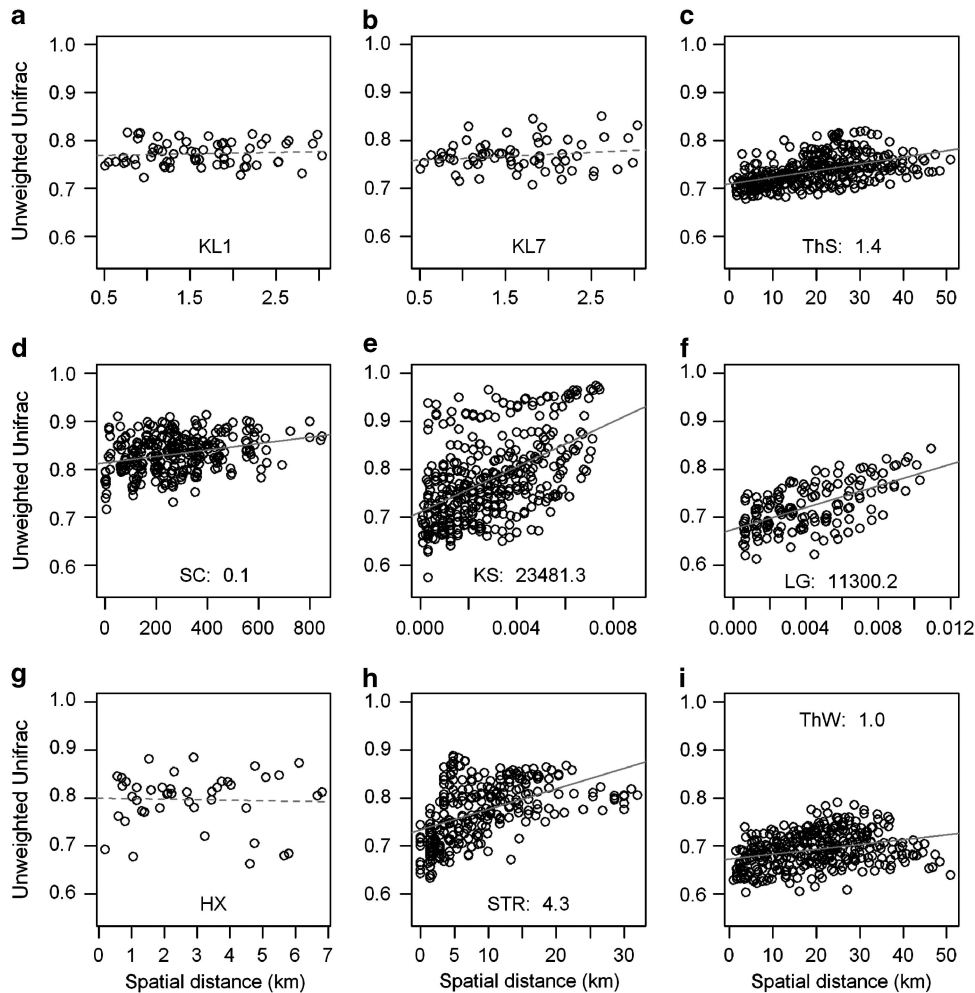
Figure 3 The relationships between unweighted Unifrac and spatial distance for different sample groups. (**a–d**) Lake surface sediments; (**e**, **f**) lake subsurface sediments; (**g**) soils; (**h**) stream biofilm and (**i**) lake water. The regression slopes of the linear relationships based on Gaussian generalized model are shown with solid (statistically significant, ranked Mantel test, 9999 permutations, $P<0.05$) or dashed (statistically nonsignificant, $P>0.05$) lines. The significant slope (unweighted Unifrac per 10^3 km) is shown in each sample group panel. Detailed Mantel statistics are shown in Supplementary Table S1.

Table 2 Mantel and partial Mantel tests for the correlation between ses.betaMNTD and the explanatory distances (elevational, geographic and environmental distance) using Spearman's rho for different habitat types and spatial scales

Ses.betaMNTD		Local							Regional
Effects of	Controlling for	Surface sediments			Lake water	Subsurface sediments		Stream biofilm	Surface sediments
		KL1	KL7	ThS	ThW	_G	KS	STR	SC
Geographic		0.151	0.202	0.353***	0.244**	0.177*	0.394***	0.376**	0.145
Environmental		0.338**	0.404*	0.406***	0.386***	0.407***	0.413**	0.410***	0.387**
Elevational								0.734***	0.157
Geographic	Environmental elevational	0.078	0.140	0.173*	0.074	−0.067	0.317***	0.240*	0.029
								−0.413*	0.125**
Elevational	Environmental geographic							0.668*	0.059
								0.745*	0.139**
Environmental	Geographic elevational	0.315*	0.380*	0.273*	0.316**	0.378***	0.341**	0.296*	0.364*
								−0.007*	0.363**

The significances are tested based on 10 000 permutations. ***$P<0.001$; **$P<0.01$; *$P<0.05$. For the HX group, both the geographic and environmental distances were not significantly related to bacterial phylogenetic turnover ($P>0.05$). For ZJ and NJ groups, the Mantel analyses are not conducted because of their lower sample numbers (7 and 5, respectively).

expectations. This pattern-to-process linkage requires phylogenetic signal in microbial habitat associations. Compared with previous microbial studies, we used a more statistically robust method to test phylogenetic signal. We used this updated method to provide the broadest evaluation of phylogenetic signal in microbes to date, and find significant phylogenetic signal across all evaluated habitats. Further, we used a novel statistical approach to test for deterministic processes governed by *unmeasured* environmental variables. The ability to detect deterministic influences by *unmeasured* environmental variables proved critical for understanding the relative balance between deterministic and stochastic processes.

Phylogenetic signal varies with phylogenetic distance
Inferring ecological processes using phylogenetic information requires phylogenetic signal (Losos, 2008) in ecological niches (Cavender-Bares et al., 2009). Recent studies on freshwater *Actinobacteria* (Newton et al., 2007), marine bacterioplankton (Andersson et al., 2010) and subsurface bacteria (Stegen et al., 2012) have indicated there is a positive relationship between phylogenetic distances and ecological differences among close relatives. Our Mantel correlogram analyses support this finding: significant phylogenetic signal was consistently detected across all studied habitat types, but only across short phylogenetic distances. This is also supported by regressing habitat differences against phylogenetic distances between pairs of OTUs (the same procedures in Stegen et al., 2012), which also showed a clear positive relationship across the short phylogenetic distances for all sample groups (Supplementary Figure S5). More quantitatively, significant phylogenetic signal was found at up to 10–30% of the maximum observed phylogenetic distance. This is consistent with Stegen et al. (2012) who found up to 13–15% of the maximum phylogenetic distance for terrestrial subsurface bacteria. This general pattern in phylogenetic signal strongly indicates that closely related bacterial taxa are ecologically coherent and that interspecies gene exchange, such as horizontal gene transfer (Popa and Dagan, 2011), does not eliminate such ecological coherence at the scale of bacterial metacommunities (see also in Philippot et al., 2010; Wiedenbeck and Cohan, 2011; Stegen et al., 2012).

Unexpectedly, across intermediate distances there were significant negative correlations between phylogenetic and ecological distances. This may suggest convergent evolution, that is, that distinctly related lineages acquire the similar ecological niches, across intermediate phylogenetic distances. We are unaware of other work showing convergent evolution across free-living bacteria, but the same genes are often lost in obligate intracellular bacteria from different phyla, suggesting evolutionary convergence (Merhej et al., 2009). More generally,

evolutionary convergence may have a role in common functions for complex symbiont communities across phylogenetically divergent hosts (Fan et al., 2012). However, the causes and consequences of convergent evolution in free-living microbial communities are unclear, but warrant further study.

Taken together, our results combined with previous studies, indicate a general pattern in the phylogenetic structure of bacterial ecological niches: conserved niches/traits across short phylogenetic distances, convergent niches/traits across intermediate distances and random niches/traits across large distances. As functional and phylogenetic beta diversity for soil microbes were closely correlated (Fierer et al., 2012), and there is a phylogenetic signal for 93% functional traits in micro-organisms (Martiny et al., 2012), it would be interesting to use other molecular markers (functional genes, for instance) to test phylogenetic signal at finer phylogenetic scales within a hierarchy of environmental factors (see Martiny et al., 2009). Nevertheless, this observation has two important implications. First, strong phylogenetic signal across short phylogenetic distances indicates that ecological processes can be inferred by studying spatial or temporal patterns in the phylogenetic structure of communities. Second, it suggests that ecological inferences are most robust when made using metrics of nearest neighbor distances (for example, beta-MNTD). These metrics focus on relatively short phylogenetic distance such that phylogenetic structure carries ecologically relevant information.

Turnover rate in community composition varies among habitats
Our results showed much greater turnover in community composition between habitats than within habitats. This suggests that bacteria are specialized on particular habitats and is consistent with former meta-analyses on bacteria (for example, Lozupone and Knight, 2007; Delmont et al., 2011; Nemergut et al., 2011; Zinger et al., 2011).

Within habitats, unweighted Unifrac and betaMNTD both showed significant distance–decay patterns across six out of the nine habitat types (67%) studied here. This is consistent with Hanson et al. (2012), who found that microbial communities showed significant spatial patterns in 68% of 54 reviewed data sets. In addition, there was substantial across-habitat variation in the rate of spatial distance–decay. As expected, the turnover rate in phylogenetic community composition was highest for shallow terrestrial subsurface environments (LG and KS). This high rate of turnover in phylogenetic community composition is consistent with a previous observation of high turnover rate in taxonomic composition in a terrestrial subsurface environment (Wang et al., 2008). Such high turnover rates may be explained by strong dispersal limitation and steep environmental gradients in subsurface environments (Wang et al., 2008).

In amphibian, bird, mammal or plant assemblages, beta diversity is typically higher in mountainous regions than in regions with less topographic relief, presumably due to species specializing on particular elevations (for example, McKnight et al., 2007). Our results are consistent with this observation: turnover rate was significantly higher for the biofilm bacterial communities in mountainside streams than for other habitat types (except subsurface environments). This high elevational turnover rate of bacteria is also consistent with the results for diatoms and macroinvertebrates in the same streams (Wang et al., 2012b). On the other hand, high turnover across elevations for bacteria is somewhat different from a previous result obtained using denaturing gradient gel electrophoresis along the same elevational gradient, which showed no significant elevational distance–decay relationship (Wang et al., 2012b). This difference may have resulted from different resolution of the two methods: a method with lower resolution may fail to detect a significant distance–decay relationship because of undetected endemism (Morlon et al., 2008; Hanson et al., 2012).

Our samples covered a wide range of horizontal spatial extents, which potentially affects the observed turnover rate. Below the spatial extent of 10 km, we did not find significant distance–decay in KL1, KL7 or HX sample groups. At a spatial extent of < 100 km, as within habitats of Taihu Lake (ThS and ThW) for instance, we found significant differences in turnover rates across-habitat types: bacterial communities in surface sediments showed a significantly higher turnover rate than their corresponding free-living communities (1.4 and 1.0 unweighted Unifrac per 10^3 km, respectively). When larger spatial extents were considered (> 100 km), the lakes from mountain regions (SC) for instance, the sediment bacterial communities showed a significantly lower turnover rate than other habitats, especially within habitats (that is, ThS) (Figure 3).

Previous work has also found the distance–decay relationship for microbes to be scale dependent, where significant relationships occurred only across local or relatively short spatial extents (for example, King et al., 2010; Martiny et al., 2011). However, for larger spatial extents similar to those we considered here, former reports indicate that the bacterial distance–decay relationships range from significant in lake surface sediments across the Tibetan Plateau (Xiong et al., 2012) to nonsignificant in North America soils (Fierer and Jackson, 2006). These results collectively suggest that the rate of distance–decay in bacteria shows strong context-dependency, potentially driven by among-habitat variation in the degree of environmental spatial autocorrelation and in the degree of dispersal limitation.

Summarily, the pairwise phylobetadiversity significantly increased with spatial distance for most of the sample groups and clearly showed that among the studied habitats there were significant differences in the rates at which community composition changes through space. In general, the horizontal turnover rates of bacterial communities from lakes or soils, or from local or regional scales, were significantly lower than the rates from subsurface environments or mountain regions. In addition, the community turnover rate in subsurface environments was highest.

Deterministic processes govern community composition across habitats

By leveraging phylogenetic information and following the three-step procedure proposed here, we inferred the relative influences of deterministic and stochastic processes across a broad range of habitat types. The first step is to examine distributions of ses.betaMNTD; a distribution mean that deviates significantly from zero suggests a strong influence of deterministic processes (Fine and Kembel, 2011; Stegen et al., 2012). In 9 of the 11 sample groups, ses.betaMNTD distributions deviated significantly from zero (Supplementary Figure S3). Furthermore, seven groups showed distributions greater than zero (Supplementary Figure S3), suggesting that across communities there are shifts in environmental conditions that deterministically cause changes in community composition. Two groups from Taihu Lake (ThS and ThW) had mean ses.betaMNTD values that were less than zero (Supplementary Figure S3), suggesting that for both groups there was relative consistency in the environmental conditions that deterministically governed community composition. Although many features of the observed abiotic environment in Taihu Lake varied across sampled communities, a high level of eutrophication was maintained across communities (Duan et al., 2009). It may therefore be that high eutrophication imposed strong environmental filtering on microbial communities in Taihu Lake. More generally, our observation of ses.betaMNTD values ranging from negative to null to positive highlights the fact that the influence of deterministic ecological processes varies across systems; deterministic processes can minimize spatial variation in, have little influence over, or drive large shifts in community composition. A major challenge for future work is to mechanistically understand variation in the influence of deterministic processes.

The second step in the process-inference procedure focuses on revealing which process is the primary cause of significant, partial Mantel coefficients relating turnover in community composition to spatial distance. A common interpretation is that such a relationship is caused by stochastic processes. Although intuitive, this interpretation may be premature, especially if turnover in community composition is quantified using a observed or 'raw' metric; a raw metric such as betaMNTD simply measures the difference in composition between two communities.

Consider a scenario in which the turnover in community composition is quantified using betaMNTD and in which there is an *unmeasured* environmental variable that changes across sampled microbial communities. If this *unmeasured* variable governs community composition, it can cause a significant partial Mantel coefficient relating betaMNTD to spatial distances. The standard (and incorrect) inference would be that community composition is governed by stochastic processes.

To make a more robust inference we use ses. betaMNTD as the turnover metric. In this case, the partial Mantel coefficient related to spatial distance should reflect deterministic processes governed by *unmeasured* environmental variables. The reason is twofold: (i) the influence of measured environmental variables has already been accounted for because we are dealing with partial Mantel coefficients; and (ii) the magnitude of phylogenetic null model departures (that is, ses.betaMNTD) should only be influenced by deterministic processes; stochastic processes should have no influence (Hardy, 2008).

When using ses.betaMNTD as the turnover metric, a significant, partial Mantel coefficient related to spatial distances should indicate that stochastic processes are overwhelmed by deterministic processes governed by *unmeasured* environmental variables. Similarly, if the partial Mantel coefficient is nonsignificant, it would indicate that *unmeasured* environmental variables have little influence over community composition.

In ThS, KS and STR groups, spatial distances were significantly related to ses.betaMNTD after controlling for measured environmental distances (Table 2). We therefore infer that in these habitats there are *unmeasured*, spatially structured environmental variables that influence community composition by imposing deterministic processes. We also infer that deterministic processes imposed by *unmeasured* variables overwhelm any influences of stochastic processes. In contrast, for the other six groups, *unmeasured* environmental variables have little influence. For these six groups, a significant influence of stochastic processes can be indicated by a significant relationship between spatial distance and Unifrac or betaMNTD (after controlling for measured environmental variables).

The third step in our process-inference procedure evaluates the relative balance between deterministic and stochastic processes. To begin, we infer a greater influence of deterministic processes for the three groups characterized by significant influences of measured and *unmeasured* environmental variables (ThS, KS and STR). This inference assumes that if stochastic processes were more influential than deterministic processes, spatial distances alone would not be related to ses.betaMNTD; stochastic processes acting alone should cause there to be no relationship between ses.betaMNTD and spatial distance.

For the six sample groups in which ses.betaMNTD was not related to spatial distances (by partial

Mantel), we use the relative magnitudes of partial Mantel coefficients from the analyses of Unifrac and betaMNTD. The reason we use these observed or 'raw' beta diversity metrics instead of ses.betaMNTD is because they can increase with an increased influence of stochastic processes. That is, increased stochasticity should increase taxonomic turnover and increased taxonomic turnover should, by itself, cause increases in phylogenetic turnover. For five of the six sample groups, the partial Mantel coefficient was larger for environmental distance than for spatial distance (Supplementary Tables S1 and S2). We take this as evidence that deterministic processes are more influential than stochastic processes in these five groups. The exception was the HX group, which was characterized by nonsignificant partial Mantel tests, which does not provide any clear ecological inferences.

It is worth noting that some inferences drawn here are opposite to those which we would have made using 'traditional' approaches without null models. When a higher partial Mantel coefficient is observed for spatial distance than for environmental distance, the standard inference is that stochastic processes have a stronger influence than deterministic processes. Using this approach (for example, for KS and STR), partial Mantel coefficients based on betaMNTD would suggest a stronger influence of stochastic processes (Supplementary Table S2).

Considering that ses.betaMNTD is significantly related to spatial distance in KS and STR, however, implies that the larger coefficient on spatial distance is actually driven by *unmeasured* environmental variables that deterministically govern community composition. This reverses the inference from the dominance of stochastic processes to the dominance of deterministic processes. We suggest that the approach used here provides more informed inferences than the standard Mantel framework. It would be informative to apply the approach to the previously studied microbial and macro-organism systems. Substantial changes in our new understanding of these systems may result. We stress, however, that both approaches are important and the relative utility of each will depend on the context and questions of interest.

Concluding remarks

Although our results suggest that bacterial communities are governed primarily by deterministic processes, stochastic processes are also important. For instance, many pairwise comparisons produced nonsignificant ses.betaMNTD values and the ses.betaMNTD distributions for the Kuilei Lake were not significantly different from zero (Supplementary Figure S3). In addition, bacterial communities in surface sediments near river mouths (islands or waterway channels) were more phylogenetically similar to the lake water than the other locations
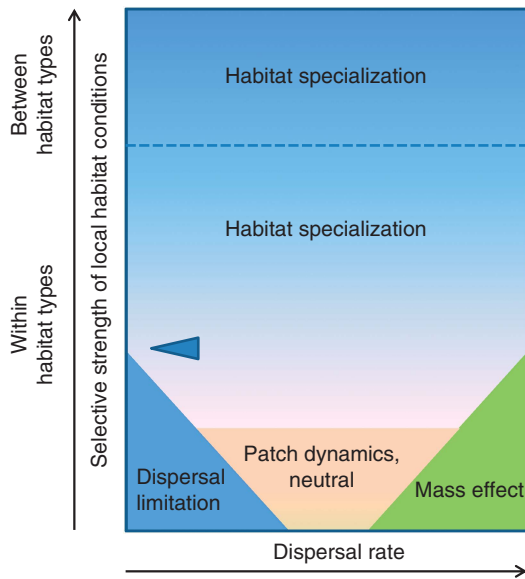
Figure 4 Conceptual figure showing the relationship between the different underlying processes for community assembly, which is adopted from Lindström and Langenheder (2012). According to Martiny *et al.* (2006) and the metacommunity framework (Leibold *et al.*, 2004), communities are mainly assembled by habitat differentiation (or species sorting), dispersal limitation, mass effects and patch dynamics/neutral processes. Different processes are assumed to act within the two dimensions of the selective strength by the local environmental conditions and the rate of dispersal among communities. The dominant roles of each process were illustrated by the colored regions within the two dimensions of dispersal rate and selective strength of local habitat conditions for new immigrating species. The conceptual threshold separating the dominant role of deterministic and stochastic processes was shown by a shaded triangle along the selective strength of local habitat conditions (refer to main text for explanation). The dashed line indicates the separation between the communities 'between-habitat types' and those 'within-habitat types'.

in Taihu Lake, with an increasing trend in dissimilarities from the river mouth to the lake center (Supplementary Figure S1). This pattern suggests a stronger influence of dispersal and mass effects near the river mouth, potentially due to physical mixing. These observations emphasize the need to understand what governs the relative balance between stochastic and deterministic processes and what conditions would lead to stochastic processes overwhelming deterministic processes.

By integrating former theoretical reviews (Leibold *et al.*, 2004; Martiny *et al.*, 2006), Lindström and Langenheder (2012) depicted the relationship between stochastic and deterministic processes along gradients of dispersal rate and selective strength imposed by local habitat conditions (Figure 4). Our results across a broad range of ecosystems suggest that the dominance of stochastic processes might occur only when the selective strength of local habitat conditions is below a conceptual threshold (Figure 4). Microbial communities are above this threshold when they are considered across-habitat types or across space within highly environmentally heterogeneous

habitats. This is likely driven by habitat specialization, reflected here by significant phylogenetic signal in environmental associations across short phylogenetic distances. In systems with less environmental variation or with regional species pools characterized by environmental generalists, stochastic processes may overwhelm deterministic processes, as suggested in previous work (for example, Ofiteru *et al.*, 2010).

Acknowledgements

References

Anderson MJ. (2001). A new method for non parametric multivariate analysis of variance. *Austral Ecol* **26**: 32–46.

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL *et al.* (2011). Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecol Lett* **14**: 19–28.

Andersson AF, Riemann L, Bertilsson S. (2010). Pyrosequencing reveals contrasting seasonal dynamics of taxa within Baltic Sea bacterioplankton communities. *ISME J* **4**: 171–181.

Besemer K, Peter H, Logue JB, Langenheder S, Lindstrom ES, Tranvik LJ *et al.* (2012). Unraveling assembly of stream biofilm communities. *ISME J* **6**: 1459–1468.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **7**: 335–336.

Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. (2009). The merging of community ecology and phylogenetic biology. *Ecol Lett* **12**: 693–715.

Chase JM, Myers JA. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Phil Trans R Soc B* **366**: 2351–2363.

Cottenie K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecol Lett* **8**: 1175–1182.

Darwin C. (1859). *On the Origin of Species*. John Murray: London, UK.

Delmont TO, Malandain C, Prestat E, Larose C, Monier J-M, Simonet P *et al.* (2011). Metagenomic mining for microbiologists. *ISME J* **5**: 1837–1843.

Diniz-Filho JAF, Terribile LC, da Cruz MJR, Vieira LCG. (2010). Hidden patterns of phylogenetic non-stationarity overwhelm comparative analyses of niche conservatism and divergence. *Glob Ecol Biogeogr* **19**: 916–926.

Duan H, Ma R, Xu X, Kong F, Zhang S, Kong W *et al.* (2009). Two-decade reconstruction of algal blooms in China's Lake Taihu. *Environ Sci Technol* **43**: 3522–3528.

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS *et al.* (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci USA* **109**: E1878–E1887.

Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* **103**: 626–631.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL *et al.* (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA* **109**: 21390–21395.

Fine PVA, Kembel SW. (2011). Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography* **34**: 552–565.

Graham CH, Fine PVA. (2008). Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecol Lett* **11**: 1265–1277.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Hardy OJ. (2008). Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J Ecol* **96**: 914–926.

Hubbell SP. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press.

King AJ, Freeman KR, McCormick KF, Lynch RC, Lozupone C, Knight R *et al.* (2010). Biogeography and habitat modelling of high-alpine bacteria. *Nat Commun* **1**: 53.

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

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.

Lindström ES, Langenheder S. (2012). Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* **4**: 1–9.

Liu X, Dong H, Yang X, Herzschuh U, Zhang E, Stuut J-BW *et al.* (2009). Late Holocene forcing of the Asian winter and summer monsoon as evidenced by proxy records from the northern Qinghai–Tibetan Plateau. *Earth Planet Sci Lett* **280**: 276–284.

Losos JB. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* **11**: 995–1003.

Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Lozupone CA, Knight R. (2007). Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* **104**: 11436–11440.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. (2009). Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ Microbiol* **11**: 823–832.

Martiny AC, Treseder K, Pusch G. (2012). Phylogenetic conservatism of functional traits in microorganisms. *ISME J;* e-pub ahead of print 13 December 2012 doi:10.1038/ismej.2012.160.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proc Natl Acad Sci USA* **108**: 7850–7854.

McKnight MW, White PS, McDonald RI, Lamoreux JF, Sechrest W, Ridgely RS *et al.* (2007). Putting beta-diversity on the map: broad-scale congruence and coincidence in the extremes. *PLoS Biol* **5**: e272.

Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* **4**: 13.

Morlon H, Chuyong G, Condit R, Hubbell S, Kenfack D, Thomas D *et al.* (2008). A general framework for the distance-decay of similarity in ecological communities. *Ecol Lett* **11**: 904–917.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al.* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135–144.

Newton RJ, Jones SE, Helmus MR, McMahon KD. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Appl Environ Microbiol* **73**: 7169–7176.

Oden NL, Sokal RR. (1986). Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Syst Biol* **35**: 608–617.

Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA *et al.* (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci USA* **107**: 15345–15350.

Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB *et al.* (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* **8**: 523–529.

Popa O, Dagan T. (2011). Trends and barriers to lateral gene transfer in prokaryotes. *Curr Opin Microbiol* **14**: 615–623.

Stegen JC, Hurlbert AH. (2011). Inferring ecological processes from taxonomic, phylogenetic and functional trait β-diversity. *PLoS One* **6**: e20906.

Stegen JC, Lin X, Konopka AE, Fredrickson JK. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653–1664.

Tuomisto H, Ruokolainen K. (2006). Analyzing or explaining beta diversity? Understanding the targets of different methods of analysis. *Ecology* **87**: 2697–2708.

Vellend M. (2010). Conceptual synthesis in community ecology. *Q Rev Biol* **85**: 183–206.

Wang J, Soininen J, He J, Shen J. (2012a). Phylogenetic clustering increases with elevation for microbes. *Environ Microbiol Rep* **4**: 217–226.

Wang J, Soininen J, Zhang Y, Wang B, Yang X, Shen J. (2011). Contrasting patterns in elevational diversity between microorganisms and macroorganisms. *J Biogeogr* **38**: 595–603.

Wang J, Soininen J, Zhang Y, Wang B, Yang X, Shen J. (2012b). Patterns of elevational beta diversity in micro- and macroorganisms. *Glob Ecol Biogeogr* **21**: 743–750.

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

Webb CO, Ackerly DD, McPeek MA, Donoghue MJ. (2002). Phylogenies and community ecology. *Annu Rev Ecol Syst* **33**: 475–505.

Wiedenbeck J, Cohan FM. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* **35**: 957–976.

Xiong J, Liu Y, Lin X, Zhang H, Zeng J, Hou J *et al.* (2012). Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ Microbiol* **14**: 2457–2466.

Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM *et al.* (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One* **6**: e24570.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)