ORIGINAL ARTICLE

Stochastic and deterministic assembly processes in subsurface microbial communities

James C Stegen, Xueju Lin[1], Allan E Konopka and James K Fredrickson

Fundamental and Computational Sciences Directorate, Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA

A major goal of microbial community ecology is to understand the forces that structure community composition. Deterministic selection by specific environmental factors is sometimes important, but in other cases stochastic or ecologically neutral processes dominate. Lacking is a unified conceptual framework aiming to understand why deterministic processes dominate in some contexts but not others. Here we work toward such a framework. By testing predictions derived from general ecological theory we aim to uncover factors that govern the relative influences of deterministic and stochastic processes. We couple spatiotemporal data on subsurface microbial communities and environmental parameters with metrics and null models of within and between community phylogenetic composition. Testing for phylogenetic signal in organismal niches showed that more closely related taxa have more similar habitat associations. Community phylogenetic analyses further showed that ecologically similar taxa coexist to a greater degree than expected by chance. Environmental filtering thus deterministically governs subsurface microbial community composition. More importantly, the influence of deterministic environmental filtering relative to stochastic factors was maximized at both ends of an environmental variation gradient. A stronger role of stochastic factors was, however, supported through analyses of phylogenetic temporal turnover. Although phylogenetic turnover was on average faster than expected, most pairwise comparisons were not themselves significantly non-random. The relative influence of deterministic environmental filtering over community dynamics was elevated, however, in the most temporally and spatially variable environments. Our results point to general rules governing the relative influences of stochastic and deterministic processes across micro- and macro-organisms.
The ISME Journal (2012) **6,** 1653–1664; doi:10.1038/ismej.2012.22; published online 29 March 2012
Subject Category: microbial population and community ecology
Keywords: community assembly; distance decay; evolutionary niche conservatism; neutral theory; niche theory; phylogenetic beta diversity

Introduction

A long-standing challenge is to understand processes governing the composition of ecological communities (Clements, 1916; Gleason, 1927; Diamond, 1975; Weiher and Keddy, 1995; Chesson, 2000; Hubbell, 2001). It is broadly recognized that community assembly is simultaneously influenced by factors that are relatively deterministic and factors that are more stochastic (for example, Dumbrell *et al.*, 2010; Ofiteru *et al.*, 2010; Langenheder and Szekely, 2011). The deterministic class includes selection imposed by the abiotic environment ('environmental filtering') and both antagonistic and synergistic species interactions.

In contrast, the stochastic class includes unpredictable disturbance, probabilistic dispersal and random birth-death events (see Chase and Myers, 2011). The most immediate goals are to uncover how and why the relative influences of stochastic and deterministic factors vary through space and time (Chase, 2007, 2010; Chase *et al.*, 2009; Myers and Harms, 2009, 2010; Chase and Myers, 2011).

Although there is a rich literature examining microbial community composition across disparate habitats, few studies examine how the relative influences of stochastic and deterministic processes change with environmental conditions (but see Horner-Devine and Bohannan, 2006; Bryant *et al.*, 2008; Kembel *et al.*, 2011). Approaches commonly applied to microbial communities generally lead to inferences regarding whether or not community composition is significantly related to specific environmental factors (for example, Fuhrman *et al.*, 2006; Sun *et al.*, 2006; Bork *et al.*, 2009; Zinger *et al.*, 2011; Gilbert *et al.*, 2012). These inferences are clearly important, but we suggest that

1654

applying the concepts and statistical tools that are more commonly used in macroorganism ecology provides a complementary approach that can significantly extend the understanding of microbial communities. We specifically couple phylogenetic community composition with randomization procedures, referred to as 'null models' (Webb, 2000; Webb et al., 2002; Graham and Fine, 2008), and examine deviations from null model expectations across temporal and spatial environmental gradients. Doing so characterizes how the relative influences of stochastic and deterministic processes change with environmental conditions (Chase and Myers, 2011; Chase et al., 2011).

Using phylogenetic information to infer ecological processes requires that phylogenetic distance between taxa approximate their ecological niche difference (for example, the difference in habitat requirements). When phylogenetic distance does approximate niche difference, niches are said to have 'phylogenetic signal' and to be 'phylogenetically structured' (Losos, 2008). Previous work suggests that closely related bacteria have similar habitat associations (Andersson et al., 2010) and that bacteria may be ecologically coherent even at the phylum level (Fierer et al., 2007, Philippot et al., 2010). On the other hand, horizontal gene transfer can move ecological attributes among bacteria (Doolittle, 1999a,b; Welch et al., 2002). Wiedenbeck and Cohan (2011) draw from Haldane's (1932) classical work to demonstrate that horizontal gene transfer does not, however, 'scramble' ecological niches among taxa (see also Cohan and Koeppel, 2008; Philippot et al., 2010). Horizontal gene transfer should instead improve the performance of recipients (Cohan and Koeppel, 2008; Wiedenbeck and Cohan, 2011). Rates of adaptive evolution may therefore increase because advantageous functions can be acquired from other taxa, instead of relying on mutation (Cohan and Koeppel, 2008). Rapid adaptive evolution suggests that bacterial niches should have phylogenetic signal only among close relatives. There should be no phylogenetic signal across deep evolutionary time because adaptive evolution fueled by horizontal gene transfer should not consistently direct all members of a given clade toward a specific niche. We test this expectation before examining phylogenetic community composition.

If across recent evolutionary time more closely related taxa are more ecologically similar, ecological theory links specific patterns of phylogenetic community composition to specific types of processes (Kraft et al., 2007; Fine and Kembel, 2011; Pei et al., 2011). First, if community structure and dynamics are primarily stochastic, community phylogenetic composition and dynamics should not differ significantly from expectations based on random community assembly. Second, if environmental filtering is the most influential process, coexisting taxa should be more closely related than expected by chance at a given point in time. Third, if environmental change deterministically shifts community composition, phylogenetic turnover between assemblages should be greater than expected after controlling for observed turnover in operational taxonomic units (OTUs) composition (referred to as 'taxonomic turnover'). Fourth, if competition is strong whereby coexisting taxa must be ecologically differentiated from each other, coexisting taxa should be more distantly related than expected by chance for a given point in time. This mechanism of competition is not, however, related to patterns of phylogenetic turnover.

We apply these pattern-to-process linkages to better understand factors governing (i) the composition of subsurface microbial communities; and (ii) changes in the relative influences of stochastic and deterministic processes. Our study further evaluates the utility of including phylogenetic information and associated randomization procedures into analyses of microbial systems. This is achieved through comparison with a companion study (Lin et al., 2012), which examines the same data set from a conceptual and statistical perspective more commonly applied to microbial systems.

Materials and methods

Data collection
A detailed description of sampling methods is provided in Lin et al. (2012). Briefly, nine wells spread across three clusters of three wells each were sampled within the Hanford Integrated Field Research Challenge site (http://ifchanford.pnnl.gov/); North of Richland, WA (Bjornstad et al., 2009). Clusters were ∼30 m apart, forming vertices of an equilateral triangle, and the three wells within each cluster were within 0.6 m of each other. Clusters A, B and C were ∼270 m, 250 m and 280 m, respectively, from the Columbia River. All wells were ∼15 cm in diameter and each well within each cluster was screened at one depth: 10, 13 or 17 m below ground surface. An extensive characterization of all environmental variables potentially relevant to bacterial physiology is not yet available for our system. However, it is known that sediment texture varies vertically (Bjornstad et al., 2009) and that the magnitude of river water intrusion varies with depth and with horizontal distance from the Columbia River (McKinley et al., in review). From December 2009 to September 2010, 354 samples were taken across 41 points in time. Sampling interval varied from biweekly to every two weeks. Water table elevation was also monitored (McKinley et al., in review; Vermeul et al., 2011). Sampled groundwater (4 l) was passed through a 0.2-µm polyethersulfone filter (Supor200, Pall Corporation, Port Washington, NY, USA). DNA was extracted from half of each filter using a PowerSoil-htp 96-well DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). PCR was conducted following

Costello *et al.* (2009) and Fierer *et al.* (2008), and pyrosequencing was carried out on a Roche FLX 454 (Roche Diagnostics, Indianapolis, IN, USA). Sequences were analyzed in QIIME (Caporaso *et al.*, 2010), using PyNast (http://pynast.sourceforge.net) and the Greengenes core set (http://greengenes.lbl.gov) for alignment and FastTree (http://www.microbesonline.org/fasttree/) to estimate a phylogeny containing all OTUs observed across all samples. OTUs were defined using a 97% sequence similarity cutoff. Each sample was rarefied to 500 sequences and, subsequently, the 1000 most abundant OTUs were retained for analyses.

Phylogenetic signal
To ask whether closely related OTUs have similar habitat associations (that is, to test for phylogenetic signal in habitat associations), the abundance-weighted mean subsurface depth and water table elevation were first found for each OTU. For example for depth, we took all records of a given OTU, recorded the depth of each record and the OTU's abundance in each record, and then found the abundance-weighted mean of depth. This is the OTU's 'niche value' for depth. The analogous procedure was used to estimate OTU niche values for water elevation. The niche values estimate the depth and the water elevation where each OTU is most abundant, thereby characterizing two axes of each OTU's niche in terms of the habitats where it occurs (for similar approaches in plant and microbial ecology, see Andersson *et al.*, 2010; Pei *et al.*, 2011; Hardy *et al.*, 2012). For each niche axis, between-OTU niche differences were related to between-OTU phylogenetic distances. To summarize major trends in this relationship, between-OTU niche differences were placed in phylogenetic distance bins and median niche difference was found in each bin. Median niche differences were then regressed against phylogenetic distance bin mid-points (similar to Andersson *et al.*, 2010). Maximum phylogenetic distance was 3.64 (arbitrary units) and a bin interval of 0.01 units was used, proving good resolution of the niche difference-to-phylogenetic distance relationship. This allowed identification of the phylogenetic distance threshold beyond which niche differences no longer increased with phylogenetic distance (Figure 3). Using all data (not medians) we then conducted Mantel tests with 999 randomizations above and below the phylogenetic distance threshold (see also Diniz-Filho *et al.*, 2010). We further compared the niche difference-to-phylogenetic distance linear regression slopes above and below the phylogenetic distance threshold to the expectations under Brownian niche evolution. For both niche axes 999 replicates of Brownian niche evolution were run across the empirical phylogeny using 'sim.char' in package 'geiger'. Variance of the evolving niche values was set equal to the empirically observed variance in niche values for each niche axis. In Brownian evolution, moving from the root to the tips of the phylogeny, niche values evolve via Brownian motion: niche values take a random walk across the niche axis, where descendants have niche values similar to their ancestor.

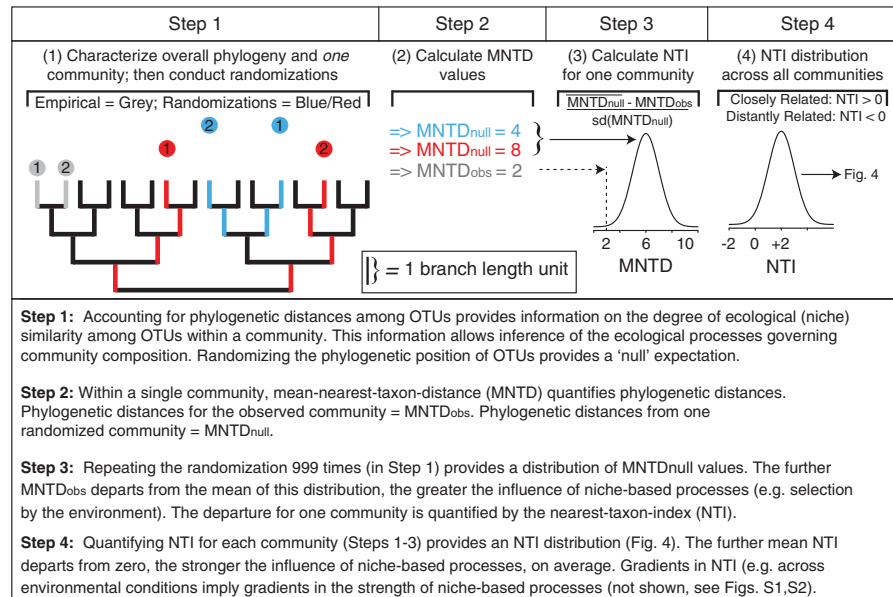*Phylogenetic community composition*
To characterize phylogenetic community composition *within* each sample (unique point in space and time), we quantified mean-nearest-taxon-distance (MNTD) and the nearest-taxon-index (NTI) (Webb *et al.*, 2002) using 'mntd' and 'ses.mntd' in package 'picante'. Note that NTI is the negative of the output of 'ses.mntd.' MNTD finds the phylogenetic distance between each OTU within a sample and its closest relative also found in the sample. The abundance weighted mean is then taken across these phylogenetic distances.

$$MNTD = \sum_{i_k=1}^{n_k} f_{i_k} \min(\Delta_{i_k j_k})$$

where f_{i_k} is the relative abundance of OTU i in community k, n_k is the number of OTUs in k, and $\min(\Delta_{i_k j_k})$ is the minimum phylogenetic distance between OTU i and all other OTUs j that are also in k.

To evaluate the degree of non-random phylogenetic community structure, OTUs and their relative abundances were randomized across the tips of phylogeny (null.model = 'taxa.labels' in 'ses.mntd'). Good Type I and Type II error rates have been found when using this randomization under conditions similar to our natural system (see Figure 3 in Kembel, 2009). NTI quantifies the number of standard deviations that the observed MNTD is from the mean of the null distribution (999 randomizations) (Figure 1). For a single community, NTI greater than $+2$ indicates coexisting taxa are more closely related than expected by chance (phylogenetic clustering). NTI less than -2 indicates coexisting taxa are more distantly related than expected by chance (phylogenetic overdispersion). A mean NTI taken across all communities that is significantly different from zero indicates clustering (NTI > 0) or overdispersion (NTI < 0) on average (Figure 1; Kembel, 2009).

Turnover in phylogenetic composition through time (phylogenetic β-diversity) was quantified using Beta Mean Nearest Taxon Distance (βMNTD) and Beta Nearest Taxon Index (βNTI), which are the between-assemblage analogs of MNTD and NTI, respectively (Figure 2; Fine and Kembel, 2011; Webb *et al.*, 2011). Instead of examining βMNTD and βNTI across all possible pairwise comparisons, we quantified these metrics for pairwise comparisons through time within each well. This was done to specifically relate temporal dynamics of community composition to the temporal dynamics of river elevation. βMNTD is the abundance-weighted-mean phylogenetic distance among closest relatives

Step 1	Step 2	Step 3	Step 4
(1) Characterize overall phylogeny and *one* community; then conduct randomizations	(2) Calculate MNTD values	(3) Calculate NTI for one community	(4) NTI distribution across all communities

Step 1: Accounting for phylogenetic distances among OTUs provides information on the degree of ecological (niche) similarity among OTUs within a community. This information allows inference of the ecological processes governing community composition. Randomizing the phylogenetic position of OTUs provides a 'null' expectation.

Step 2: Within a single community, mean-nearest-taxon-distance (MNTD) quantifies phylogenetic distances. Phylogenetic distances for the observed community = $MNTD_{obs}$. Phylogenetic distances from one randomized community = $MNTD_{null}$.

Step 3: Repeating the randomization 999 times (in Step 1) provides a distribution of $MNTD_{null}$ values. The further $MNTD_{obs}$ departs from the mean of this distribution, the greater the influence of niche-based processes (e.g. selection by the environment). The departure for one community is quantified by the nearest-taxon-index (NTI).

Step 4: Quantifying NTI for each community (Steps 1-3) provides an NTI distribution (Fig. 4). The further mean NTI departs from zero, the stronger the influence of niche-based processes, on average. Gradients in NTI (e.g. across environmental conditions imply gradients in the strength of niche-based processes (not shown, see Figs. S1,S2).

Figure 1 Summary of methods characterizing *within* community phylogenetic composition. After estimating a phylogeny that includes all OTUs across all sampled communities, only the OTUs within a given community are retained (observed OTUs shown as gray circles; numbers are OTU identities). In the example, there are two observed OTUs in the community and the phylogenetic distance between them is two branch length units (sum of gray branches). Abundances are not shown, but are assumed equal so that $MNTD_{obs}$ is 2 units. To generate a null expectation for MNTD the observed OTUs and their relative abundances are randomly placed on the phylogeny (red circles; note that OTU identities are the same, only their positions on the phylogeny have changed). MNTD is then re-calculated, providing one null MNTD value ($MNTD_{null}$). Red branches connect the two (randomly placed) OTUs and their sum is 8, so $MNTD_{null} = 8$ units. The randomization is repeated (blue circles), providing a second $MNTD_{null}$ value (4 units), and so on until 999 randomizations are completed. This provides a distribution of $MNTD_{null}$ values to which $MNTD_{obs}$ is compared (Step 3). In the example, $MNTD_{obs}$ (vertical dashed line) is much smaller than the average $MNTD_{null}$ value. The equation for the NTI is given under Step 3 and measures the difference between $MNTD_{obs}$ and mean $MNTD_{null}$ (given as $\overline{MNTD_{null}}$) in units of standard deviations, where standard deviation is measured on the $MNTD_{null}$ distribution. Repeating Steps 1–3 across all communities provides the NTI distribution in Step 4 (actual distribution shown in Figure 4). An NTI distribution with a mean greater (less) than zero indicates niche-based processes cause OTUs to be, on average, more closely (distantly) related than expected under random community assembly.

occurring in two different communities (Figure 2).

$$\beta MNTD = 0.5 \left[\sum_{i_k = 1}^{n_k} f_{i_k} \min\left(\Delta_{i_k j_m}\right) + \sum_{i_m = 1}^{n_m} f_{i_m} \min\left(\Delta_{i_m j_k}\right) \right]$$

where $\min(\Delta_{i_k j_m})$ is the minimum phylogenetic distance between OTU i in community k and all OTUs j in community m (see above for other variables). A null distribution of βMNTD is found by randomizing OTUs across the phylogeny and recalculating βMNTD 999 times. βNTI is the number of standard deviations that the observed βMNTD is from the mean of the null distribution. $\beta NTI < -2$ or $> +2$ indicates less than or greater than expected phylogenetic turnover, respectively, for one pairwise comparison. A mean βNTI taken across all pairwise comparisons that is significantly different from zero indicates greater ($\beta NTI > 0$) or less ($\beta NTI < 0$) than expected turnover on average (Figure 2; Fine and Kembel, 2011; Webb *et al.*, 2011).

Importantly, the randomizations for βNTI control for observed taxonomic turnover. High levels of taxonomic turnover (for example, large Bray–Curtis dissimilarity) may be associated with less than expected, greater than expected, or as expected (random) levels of phylogenetic turnover. For a single pairwise comparison, $\beta NTI > +2$, therefore indicates observed phylogenetic turnover is above what is expected given observed taxonomic turnover.

Parametric and non-parametric tests were used to ask whether the mean (or median) of NTI and βNTI distributions were significantly different from zero. To ask whether NTI showed systematic trends through time or with river water elevation we examined linear regressions of NTI against Julian day and against water elevation. Significant regressions would imply systematic changes in the influence of niche-based processes, but non-significant regressions do not imply static community composition: there can be significant turnover in phylogenetic composition (large βNTI) with no change in NTI. To evaluate the influence of changes in river water elevation on phylogenetic turnover, we regressed βMNTD and βNTI against temporal differences in river elevation. Two sets of regression analyses were applied to the βMNTD and βNTI data. The first set examined data across all comparisons within a well. The second examined only comparisons between samples taken in mid-May and each subsequent sample until early-July, which is the time period before and during river water

Step 1	Step 2	Step 3	Step 4
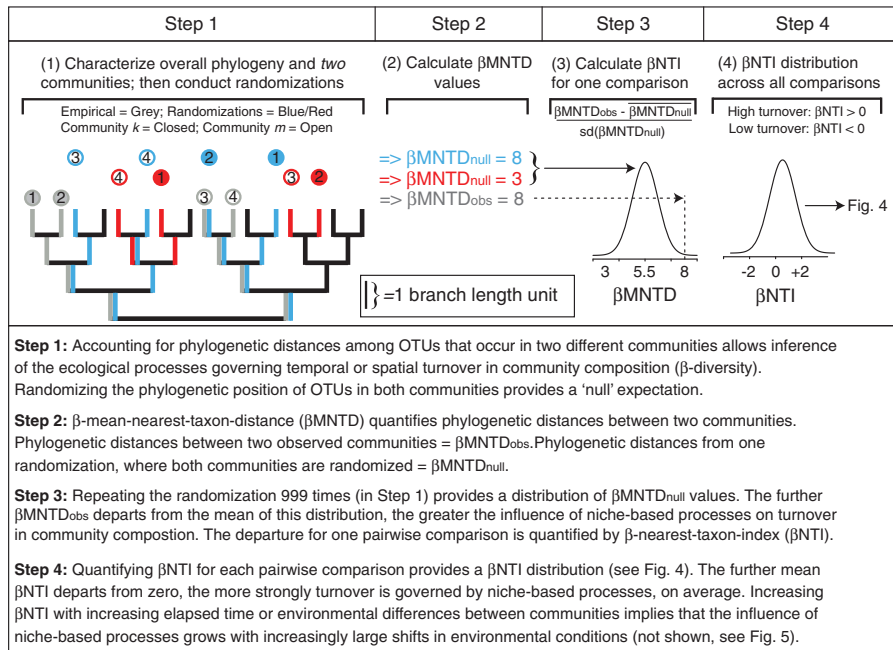
Figure 2 Summary of methods for characterizing turnover in community phylogenetic composition *between* a given pair of communities. After estimating a phylogeny that includes all OTUs across all sampled communities, only the OTUs within a given pair of communities (*k* and *m*) are retained (observed OTUs shown as gray circles; numbers are OTU identities). In the example there are two OTUs in each community. The mean of (gray) branch lengths connecting each OTU in community *k* with its closest relative in community *m* (and vice versa) gives $\beta MNTD_{obs} = 8$ branch length units (abundances assumed equal). To generate a null expectation for $\beta MNTD$ the observed OTUs are randomly placed on the phylogeny (red circles) and $\beta MNTD$ is re-calculated, always using minimum phylogenetic distances (red branches) connecting OTUs between the two communities. In the example OTU1 and OTU4 are closest relatives (4 units apart) and OTU3 and OTU2 are closest relatives (2 units apart). Taking the mean gives $\beta MNTD_{null} = 3$ units. The randomization is repeated (blue circles), providing a second $MNTD_{null}$ value (8 units, using blue branches connecting closest relatives between communities), and so on until 999 randomizations are completed. This provides a distribution of $\beta MNTD_{null}$ values to which $\beta MNTD_{obs}$ is compared. In the example, $\beta MNTD_{obs}$ (vertical dashed line) is larger than the average $\beta MNTD_{null}$ value. βNTI measures the difference between $\beta MNTD_{obs}$ and mean $\beta MNTD_{null}$ (given as $\overline{\beta MNTD_{null}}$) in units of standard deviations (equation shown in Step 3). Repeating Steps 1–3 across all pairwise comparisons provides the βNTI distribution in Step 4 (actual distribution shown in Figure 4). A βNTI distribution with a mean greater (less) than zero indicates that niche-based processes cause phylogenetic turnover between communities to be, on average, greater (less) than expected under random community assembly.

intrusion into the sampled well field (Lin *et al.*, 2012). For both sets, three statistical models were fit (1-exponential, logarithmic, and linear) using maximum likelihood, and the model with the smallest Bayesian Information Criterion (Schwarz, 1978) was chosen as the best model. Although our data come from time series, characterizations of community structure and turnover in community structure are quantified as deviations from a null model. The assumption of statistical independence is therefore upheld.

Previous studies use other metrics such as mean pairwise distance (Webb *et al.*, 2002; Webb *et al.*, 2011), Unifrac (Knight and Lozupone, 2005), and phylosor (Bryant *et al.*, 2008). These metrics account for deep evolutionary relationships among taxa, and are relevant when there is phylogenetic signal in OTU niches through both recent and deep evolutionary time. In our system phylogenetic signal occurs over relatively short phylogenetic distances (see Results). By quantifying phylogenetic distances only between nearest relatives,

the $(\beta)MNTD$ and $\overline{(\beta)NTI}$ metrics provide the most appropriate ecological inferences for our system.

Results

Regressing between-OTU niche differences against between-OTU phylogenetic distances showed a steep positive relationship across relatively short phylogenetic distances ($\sim 13–15\%$ of the maximum phylogenetic distance across the entire tree), but no systematic relationship thereafter (Figure 3). Mantel tests showed that across short phylogenetic distances ($< 13\%$ of maximum), habitat differences increased significantly with phylogenetic distances for both depth ($r = 0.14$; $P < 0.002$) and river elevation ($r = 0.12$; $P < 0.002$). The rates (i.e., linear regression slopes) at which habitat differences increased with phylogenetic distances followed the Brownian evolution expectation ($P \geqslant 0.18$ for both habitats). Across longer phylogenetic distances ($> 13\%$ of maximum) the rates at which niche
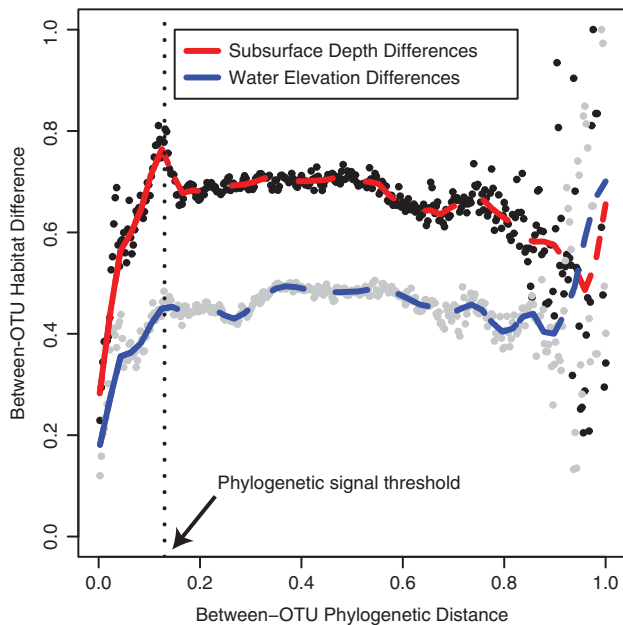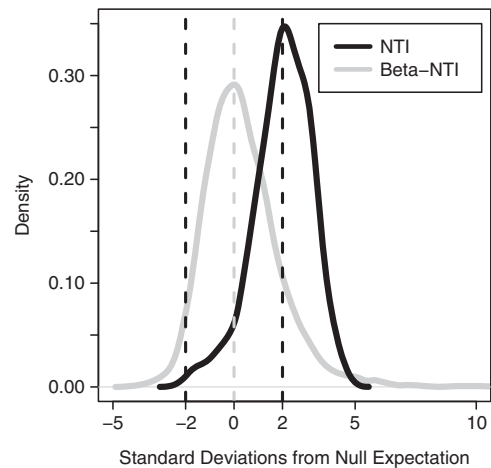
Figure 3 Median habitat differences between pairs of OTUs as a function of between OTU phylogenetic distances. Medians are taken within phylogenetic distance bins. Black and gray data points represent differences in OTU mean depth and mean river elevation, respectively. The vertical dotted line and the transition from solid to dashed curves indicates the point at which the curve fit becomes non-monotonic, thereby approximating the phylogenetic distance threshold where phylogenetic signal is lost (~13% of the maximum phylogenetic distance across the entire tree).

Figure 4 Kernal density estimates for distributions of abundance weighted nearest taxon index (NTI) and its between-community analog (βNTI). Each observation is the number of null model standard deviations the observed value is from the mean of its associated null distribution. Positive (negative) NTI values indicate shorter (longer) nearest taxon distances within a community than expected by chance. Positive (negative) βNTI values indicate greater (less) than expected turnover in phylogenetic composition than expected given the observed degree of taxonomic turnover. For both metrics, individual values below -2 or above $+2$ are statistically significant, as indicated by the dashed black lines. The dashed gray line is centered on zero, the expectation under neutral community assembly.

differences changed with phylogenetic distances were much slower than the Brownian evolution expectation ($P < 0.002$ for both habitats).

NTI across all local assemblages was significantly different from zero ($t_{353} = -31.7$, $P \ll 0.0001$), with a mean of $+2$ (Figure 4). βNTI showed a skewed distribution with a long tail stretching beyond the threshold value of $+2$, and a median of 0.16 that was significantly different from zero (Wilcoxon signed-rank test, V = 13711837, $P \ll 0.0001$) (Figure 4).

Testing for depth and well cluster effects with analysis of variance (ANOVA) showed that mean NTI varied across well clusters ($F_{2,349} = 8.2$, $P = 0.0003$) but not across depths ($F_{2,349} = 2.2$, $P = 0.11$). Mean NTI values for well clusters A, B and C were, $+1.81$, $+1.84$ and $+2.36$, respectively, and all were significantly greater than zero ($P \ll 0.0001$ for all). Note that depth and well cluster are used as categorical variables.

Distributions of βNTI within sites were skewed, and a robust non-parametric alternative is not available for two-way ANOVA with multiple samples from each treatment. As such, we used the non-parametric alternative to one-way ANOVA (that is, the Kruskal–Wallis test) to test for differences in βNTI medians across depths and well clusters. βNTI medians differed significantly across well clusters and depths ($P \ll 0.0001$ for both), with the largest medians consistently observed at the shallowest (10 m) sampling depth (Table 1).

Within most sites (one cluster, one depth), NTI did not vary with river water elevation or through time (Supplementary Figures S1 and S2). Notably, however, at the shallow depth within cluster A (see Figure 1 in Lin *et al.*, 2012) the degree of phylogenetic clustering increased with water elevation and through time (Supplementary Figures S1A and S2A).

Relating βMNTD and βNTI to change in river elevation between sample pairs revealed generally weak relationships, although some were statistically significant (Figure 5, Supplementary Figures S3 and S4). Examining comparisons between samples taken in mid-May to subsequent samples up to early-July revealed relatively strong or weak relationships between river elevation and, respectively, βMNTD or βNTI (Figure 5, Supplementary Figures S3 and S4). One important exception for the May-to-July comparisons was the shallow depth within cluster A. In this site, increasing differences in river elevation were associated with very rapid increases in βMNTD (Figure 5a) and an increase followed by saturation of βNTI above $+2$ (Figure 5d).

Discussion

Overall, we found that environmental filtering strongly determines local community composition of subsurface microbes, and that stochastic
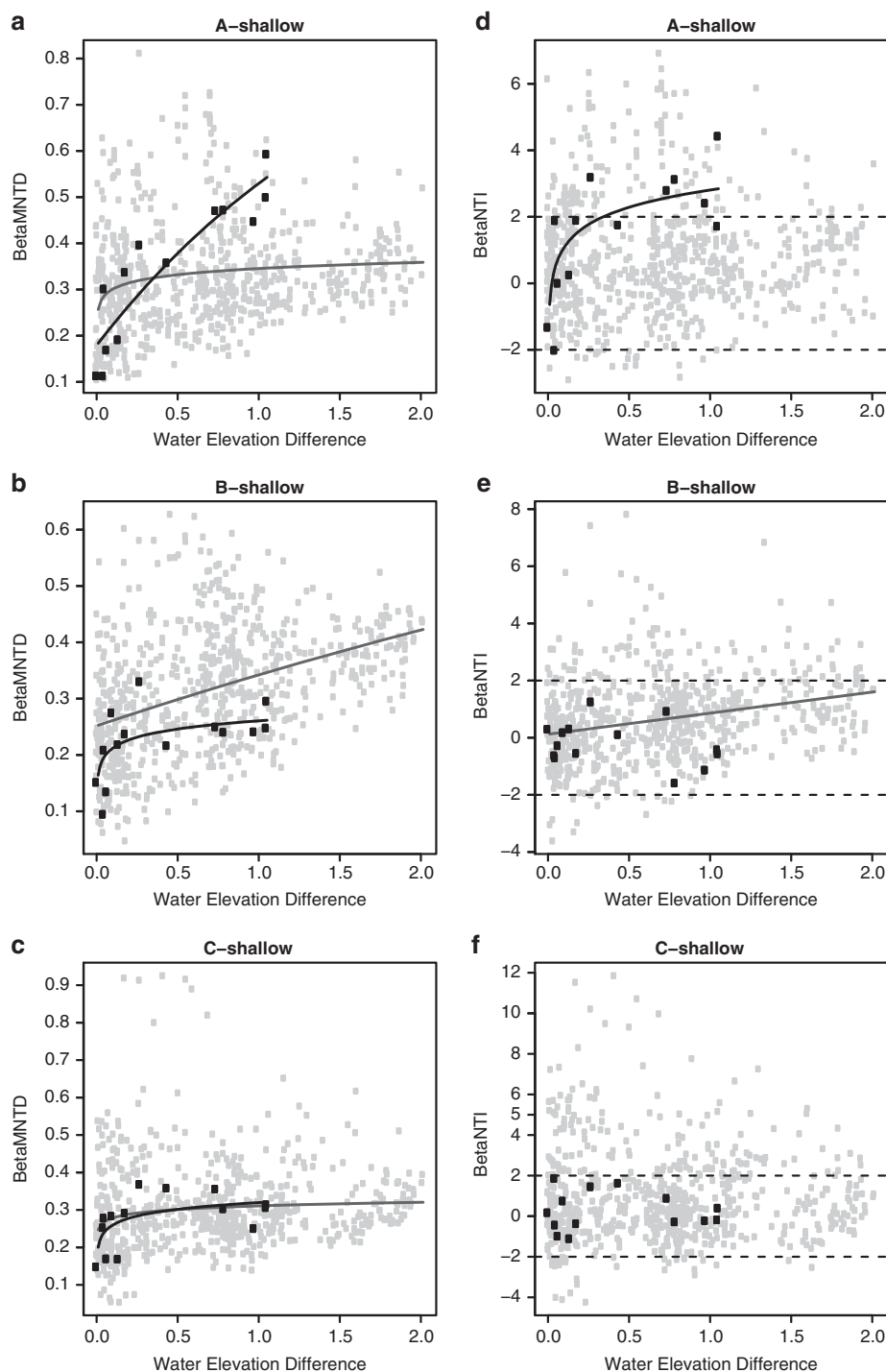
Figure 5 (**a–c**) Beta Mean Nearest Taxon Distance (βMNTD) or (**d–f**) Beta Nearest Taxon Index (βNTI) as a function of difference in Columbian River water elevation. Each panel shows data for the shallow sampling depth (10 m) across one of the three well clusters (see panel headers). Gray squares represent all comparisons from a given well cluster, and solid gray lines show best-fit regression models (linear, exponential or logarithmic). Solid black squares show results from comparing a sample in each well from mid-May to later sampling dates until early-July, and solid black lines are best-fit regression models to those data. Only significant regressions are shown. Horizontal dashed lines in (**d–f**) indicate βNTI values of − 2 and + 2, the values beyond which an individual βNTI value is considered statistically significant.

factors have an important but secondary role. The influence of stochastic factors over changes in community composition is, in comparison, much greater. However, the influence of stochastic processes relative to deterministic processes varies through space and time (see also Caruso *et al.*, 2011). Our results further suggest that general rules may govern spatial and temporal changes in the relative influences of stochastic and deterministic processes.

1660

Phylogenetic signal in ecological niches
Using phylogenetic information to make ecological inferences required that we first test for phylogenetic signal in OTU niches. Drawing on established evolutionary theory (specifically Haldane, 1932; Wiedenbeck and Cohan, 2011), we predicted that a positive relationship between bacterial ecological differences and phylogenetic distances should only occur among close relatives. Our results and previous work in bacterial systems (Andersson et al., 2010) are consistent with this expectation: beyond $\sim$13–15% of the maximum phylogenetic distance observed between any two OTUs, there was no systematic relationship between ecological difference and phylogenetic distance (Figure 3). In addition, across close relatives, the rate at which ecological differences increased with phylogenetic distances was consistent with Brownian niche evolution. Across more distant relatives, however, there was effectively no relationship between ecological differences and phylogenetic distances. Ecological niches of bacteria have therefore evolved in a Brownian-like manner across recent evolutionary time, but in a phylogenetically random manner across deeper evolutionary time (see also Andersson et al., 2010; Diniz-Filho et al., 2010; Hardy et al., 2012). Kembel (2009) found that phylogenetic community composition provides robust inferences of ecological processes for systems like ours, where (i) phylogenetic signal is consistent with Brownian evolution; (ii) the environment is heterogeneous; and (iii) more than one aspect of organismal physiology determines OTU relative performance across environmental conditions, which we assume to be true. We therefore suggest that metrics quantifying phylogenetic distances among close relatives, such as NTI, provide information useful for understanding processes that assemble bacterial communities.

Processes governing local community composition
To infer the relative influences of stochastic and deterministic processes over local community composition we coupled the observation of phylogenetic signal in habitat associations with patterns of NTI. The term 'local' refers to a sample from one point in time and space, in contrast to turnover in community composition, which is discussed in the next section. Mean NTI across all communities was $+2$ and was significantly greater than zero. This provides good evidence for a dominant role of deterministic environmental filtering (Kembel, 2009). The key inference is that one or more environmental variables limit community membership whereby closely related and ecologically similar taxa coexist to a greater degree than expected if stochastic factors such as random birth-death events and ecological drift governed community composition (Kraft et al., 2007; Fine and Kembel, 2011; Pei et al., 2011). This inference is consistent

with previous work on microbial communities in general and subsurface systems in particular (Horner-Devine and Bohannan, 2006; Bryant et al., 2008; Vamosi et al., 2009; Dumbrell et al., 2010; Herrera et al., 2010; Jones and Hallin, 2010; Amaral-Zettler et al., 2011).

Although environmental filtering is a dominant process on average, phylogenetic composition of some communities was statistically random and the relative influence of environmental filtering varied spatially. More specifically, 32% of communities had phylogenetic composition that did not differ significantly from the null expectation ($-2 < NTI < +2$). Stochastic or ecologically neutral factors therefore have an important, but secondary role in the communities studied here. NTI distributions also differed significantly across well groups: well groups A, B and C had mean NTI values of $+1.81$, $+1.84$ and $+2.36$, respectively. In addition, 63%, 67% and 73% of individual communities were significantly phylogenetically clustered (NTI $> +2$) in groups A, B and C, respectively. These results suggest that the relative influence of environmental filtering over local community composition varies through space and is strongest in group C. In addition, there was generally no relationship between mean NTI and depth, sampling date or river elevation. These patterns collectively suggest it is environmental factors varying horizontally, as opposed to vertically or temporally, that lead to changes in the relative influence of environmental filtering. An important exception was the shallow depth within group A, where phylogenetic composition transitioned from random to significantly clustered with increasing river elevation and through time.

Why is environmental filtering the strongest in group C, and why does its influence increase with river elevation and through time only in the shallow depth of group A? One explanation is that these two locations represent opposite extremes along a gradient of environmental variation through time and space. River water intrusion is a potentially important environmental driver, and the degree of intrusion varies across wells and through time due to changes in river elevation (McKinley et al., in review). Of the three well groups examined here, group C is the furthest from the river and receives the least amount of river water during intrusion events (Lin et al., 2012; McKinley et al., in review). Relative to group C, the shallow depth within group A is environmentally variable through time and space: The shallow-A well receives more river water during intrusion events (Lin et al., 2012) and is more geologically heterogeneous. In terms of geology, the shallow-A well is partially screened within a 'rip-up clast' (Bjornstad et al., 2009), which is a meters-thick piece of fine-grained, relatively impermeable material related to the Ringold formation that otherwise underlies the more permeable Hanford

formation (in which our sampling locations lie). Rip-up clasts are physically and chemically distinct 'islands' within the Hanford formation (Bjornstad *et al.*, 2009). Their presence creates a local environment that is more structurally and potentially more chemically heterogeneous than the surrounding Hanford formation.

Owing to its distance from the river, group C may receive relatively small and infrequent river-water-supplied inputs of limiting resources such as organic carbon and nutrients. It may therefore represent a relatively extreme, low resource environment. Group C was also characterized by the strongest signal of deterministic processes governing local community composition, and previous work also shows greater determinism in extreme or low resource environments (Chase, 2007, 2010; Gerisch *et al.*, 2011; Horner-Devine and Bohannan, 2006; Kembel *et al.*, 2011; van der Plas *et al.*, 2012). Stronger deterministic processes in more extreme environments may therefore be a general phenomenon across both micro- and macro-organism systems (see also Weiher *et al.*, 2011).

At the other end of the 'variation continuum', unique microbial and geological aspects of the shallow-A well point to important interactions between river water intrusion and geological heterogeneity. In particular, intruding river water may significantly change the resource environment experienced by microbial communities, and this change may be especially large in conjunction with unique rip-up clast geochemistry. The lack of change in mean NTI with river elevation within group B, which received more river water than group A (Lin *et al.*, 2012) but is not within a rip-up clast (Bjornstad *et al.*, 2009), is consistent with this hypothesis. Additional characterization of the physical and chemical environment presented by rip-up clasts will be necessary to directly evaluate their influence over microbial communities. More generally, it appears that the relative influence of environmental filtering can increase toward environmental extremes (group C) *and* in response to shifting environmental conditions (shallow-A well).

Processes governing community dynamics
An important influence of environmental filtering over temporal community dynamics was indicated by a skewed βNTI distribution with a long tail extending beyond the $+2$ significance threshold. In addition, the median of this distribution was significantly greater than zero, suggesting that on average environmental filtering deterministically governs community dynamics (for a similar approach see Fine and Kembel, 2011; Kembel *et al.*, 2011). It is interesting to note that across all comparisons 84% were consistent with random ($-2 < \beta\text{NTI} < +2$) phylogenetic turnover. This is a much larger fraction than for NTI (32% consistent

with random assembly). Together these results suggest that although, on average, deterministic factors govern composition within *and* turnover between communities, stochastic ecological drift has a stronger influence on turnover, relative to its influence on within community composition.

Uncovering the environmental factors causing non-random community dynamics is a key challenge, and our analyses provide some insight. Our analyses of phylogenetic community dynamics suggest that the relative influence of environmental filtering over community dynamics (i) is generally stronger in the shallow subsurface and (ii) can increase through time in response to environmental change in geologically heterogeneous regions. Specifically, the βNTI median was largest in the shallow sites across all three well groups (Table 1), where environmental variation was greatest (Lin *et al.*, 2012). Higher levels of temporal environmental variation may therefore deterministically drive community dynamics, consistent with recent studies linking cyclical environmental conditions to cyclical patterns in microbial community composition (Fuhrman *et al.*, 2006; Gilbert *et al.*, 2012). In the shallow-A well βNTI increased to above the significance threshold of $+2$ with increasing differences in river elevation as the river rose from mid-May to early-July. In all other sites βNTI either did not vary with river elevation or did not increase to above $+2$, suggesting a unique environmental feature of the shallow-A well. As discussed above, the shallow-A well is indeed geologically unique in that it is partially within a rip-up clast (Bjornstad *et al.*, 2009). The unique geology and βNTI patterns found within the shallow-A well point to an important interaction between intruding river water and geological heterogeneity. That is, river water intrusion into geologically heterogeneous regions of the subsurface may alter the abiotic environment in ways that impose strong directional selection (sensu Vellend, 2010) on microbial community composition.

The βNTI patterns found here complement and extend patterns found in Lin *et al.* (2012). Most importantly, Lin *et al.* uncovered large temporal changes in community composition, and found that changes in the relative abundance of specific taxa were especially marked in the shallow-A well (see their Supplementary Figure S6). These dynamics could be driven by deterministic processes

Table 1 Medians for Beta Nearest Taxon Index (βNTI) distributions from each depth within each well cluster.

	A-Cluster	*B-Cluster*	*C-Cluster*
Shallow (10 m)	0.674	0.589	0.354
Intermediate (13 m)	0.243	0.120	-0.035
Deep (17 m)	-0.490	0.124	0.166

Cluster and depth effects are significant. Note consistently larger values at the shallow depth.

associated with river water intrusion, but may also arise from random fluctuations in population sizes due to stochastic factors. The phylogenetic analyses carried out here provide an additional layer of support for the hypothesis that the river intrusion deterministically shifted community composition. However, this was true only for the shallow-A well, suggesting that in all other sites the deterministic influence of river water intrusion did not overwhelm the influence of stochastic factors. This result highlights the fact that deterministic and stochastic processes simultaneously influence community assembly, that their relative influences vary through space and time, and that using phylogenetic tools in microbial community ecology can provide insights not accessible via standard methods.

In summary, understanding gradients in the identity and strength of ecological processes is at the forefront of current ecological research (Chase and Myers, 2011), yet this paradigm has not been broadly embraced in microbial ecology. We argue that simultaneously addressing similar questions and concepts in both microbial and 'macrobial' systems will lead to a more general and unified theory of ecological communities. Here we work toward such unification by asking questions and applying statistical tools more common in studies of macro-organisms. Our results collectively suggest that the relative roles of deterministic and stochastic processes follow patterns that may generalize across systems. Further work with both micro- and macro-organisms aimed at understanding what drives gradients in the relative influences of stochastic and deterministic processes will be critical for ultimately predicting the response of communities to environmental change.

Acknowledgements

References

Amaral-Zettler LA, Zettler ER, Theroux SM, Palacios C, Aguilera A, Amils R. (2011). Microbial community structure across the tree of life in the extreme Rio Tinto. *ISME J* **5**: 42–50.

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

Bjornstad BN, Horner JA, Vermeul VR, Lanigan DC, Thorne PD. (2009). *Borehole Completion and Conceptual Hydrogeologic Model for the IFRC Well Field, 300 Area, Hanford Site*. PNNL-18340, Pacific Northwest National Laboratory: Richland, WA.

Bork P, Gianoulis TA, Raes J, Patel PV, Bjornson R, Korbel JO *et al.* (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc Natl Acad Sci USA* **106**: 1374–1379.

Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ, Green JL. (2008). Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proc Natl Acad Sci USA* **105**: 11505–11511.

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 Methods* **7**: 335–336.

Caruso T, Chan YK, Lacap DC, Lau MCY, Mckay CP, Pointing SB. (2011). Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *ISME J* **5**: 1406–1413.

Chase JM. (2007). Drought mediates the importance of stochastic community assembly. *Proc Natl Acad Sci* **104**: 17430–17434.

Chase JM. (2010). Stochastic community assembly causes higher biodiversity in more productive environments. *Science* **328**: 1388–1391.

Chase JM, Biro EG, Ryberg WA, Smith KG. (2009). Predators temper the relative importance of stochastic processes in the assembly of prey metacommunities. *Ecol Lett* **12**: 1210–1218.

Chase JM, Kraft NJB, Smith KG, Vellend M, Inouye BD. (2011). Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* **2**: Art24.

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

Chesson P. (2000). Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* **31**: 343–366.

Clements FE. (1916). *Plant Succession*. Carnegie Institution of Washington: Washington, DC.

Cohan FM, Koeppel AF. (2008). The origins of ecological diversity in prokaryotes. *Curr Biol* **18**: R1024–U1017.

Costello EK, Halloy SRP, Reed SC, Sowell P, Schmidt SK. (2009). Fumarole-supported Islands of biodiversity within a hyperarid, high-elevation landscape on Socompa Volcano, Puna de Atacama, Andes. *Appl Environ Microb* **75**: 735–747.

Diamond JM. (1975). Assembly of species communities. In: Cody ML, Diamond JM (eds) *Ecol Evol Communities*. Harvard University Press: Cambridge, MA.

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

Doolittle WF. (1999a). Lateral genomics. *Trends Biochem Sci* **24**: M5–M8.

Doolittle WF. (1999b). Phylogenetic classification and the universal tree. *Science* **284**: 2124–2128.

Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. (2010). Relative roles of niche and neutral processes in

structuring a soil microbial community. *ISME J* **4**: 337–345.

Fierer N, Bradford MA, Jackson RB. (2007). Toward an ecological classification of soil bacteria. *Ecology* **88**: 1354–1364.

Fierer N, Hamady M, Lauber CL, Knight R. (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci USA* **105**: 17994–17999.

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.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103**: 13104–13109.

Gerisch M, Agostinelli V, Henle K, Dziock F. (2011). More species, but all do the same: contrasting effects of flood disturbance on ground beetle functional and species diversity. *Oikos* 121: 508–515.

Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B *et al.* (2012). Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298–308.

Gleason HA. (1927). Further views on the succession concept. *Ecology* **8**: 299–326.

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

Haldane JBS. (1932). *The Causes of Evolution*. Longmans, Green and Company: London, UK.

Hardy OJ, Couteron P, Munoz F, Ramesh BR, Pélissier R. (2012). Phylogenetic turnover in tropical tree communities: impact of environmental filtering, biogeography and mesoclimatic niche conservatism. *Global Ecol Biogeogr*; doi:10.1111/j.1466-8238.2011.00742.x.

Herrera CM, Canto A, Pozo MI, Bazaga P. (2010). Inhospitable sweetness: nectar filtering of pollinator-borne inocula leads to impoverished, phylogenetically clustered yeast communities. *Proc R Soc B-Biol Sci* **277**: 747–754.

Horner-Devine MC, Bohannan BJM. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**: 100–108.

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

Jones CM, Hallin S. (2010). Ecological and evolutionary factors underlying global and local assembly of denitrifier communities. *ISME J* **4**: 633–641.

Kembel SW. (2009). Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol Lett* **12**: 949–960.

Kembel SW, Eisen JA, Pollard KS, Green JL. (2011). The phylogenetic diversity of metagenomes. *Plos One* **6**: e23214.

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

Kraft NJB, Cornwell WK, Webb CO, Ackerly DD. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *Am Nat* **170**: 271–283.

Langenheder S, Szekely AJ. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J* **5**: 1086–1094.

Lin X, Mckinley J, Resch CT, Lauber C, Fredrickson J, Konopka AE. (2012). Spatial and temporal dynamics of microbial community in the Hanford unconfined aquifer. *ISME J*; e-pub ahead of print 29 March 2012; doi: 10.1038/ismej.2012.26.

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

McKinley JP, Zachara JM, Resch CT, Kaluzny RM, Miller MD, Vermeul VR *et al.* (in review). River water intrusion and contaminant uranium contributions from the vadose zone to groundwater during the annual Spring rise in Columbia River stage at the Hanford Site, Washington. *Environ Sci Technol.*

Myers JA, Harms KE. (2009). Local immigration, competition from dominant guilds, and the ecological assembly of high-diversity pine savannas. *Ecology* **90**: 2745–2754.

Myers JA, Harms KE. (2010). Seed arrival and ecological filters interact to assemble high-diversity plant communities. *Ecology* **92**: 676–686.

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.

Pei NC, Lian JY, Erickson DL, Swenson NG, Kress WJ, Ye WH *et al.* (2011). Exploring tree-habitat associations in a chinese subtropical forest plot using a molecular phylogeny generated from DNA Barcode Loci. *Plos One* **6**: e21273.

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 Micro* **8**: 523–529.

Schwarz G. (1978). Estimating the dimension of a model. *Annals Stat* **6**: 461–464.

Sun FZ, Ruan QS, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA. (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**: 2532–2538.

Vamosi SM, Heard SB, Vamosi JC, Webb CO. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol Ecol* **18**: 572–592.

van der Plas F, Anderson TM, Olff H. (2012). Trait similarity patterns within grass and grasshopper communities: multitrophic community assembly at work. *Ecology*; doi:10.1890/11-0975.1.

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

Vermeul VR, McKinley JP, Newcomer DR, Mackley RD, Zachara JM. (2011). River-induced flow dynamics in long-screen wells and impact on aqueous samples. *Ground Water* **49**: 515–524.

Webb CO. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *Am Nat* **156**: 145–155.

Webb CO, Ackerly DD, Kembel S. (2011). Phylocom: Software for the analysis of phylogenetic community structure and character evolution (with phylomatic and ecoevolve). User's manual, version 4.2. http://www.phylodiversity.net/phylocom/.

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

1664

Weiher E, Freund D, Bunton T, Stefanski A, Lee T, Bentivenga S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philos Trans R Soc B: Biol Sci* **366**: 2403–2413.

Weiher E, Keddy P. (1995). Assembly rules, null models, and trait dispersion: New questions from old patterns. *Oikos* **74**: 159–164.

Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D *et al.* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci USA* **99**: 17020–17024.

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.

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 the paper on The ISME Journal website (http://www.nature.com/ismej)