Ecology, 87(7) Supplement, 2006, pp. S100–S108

PHYLOGENETIC CLUSTERING AND OVERDISPERSION IN BACTERIAL COMMUNITIES

M. Claire Horner-Devine[1,3] and Brendan J. M. Bohannan[2]

[1]*School of Aquatic and Fishery Science, University of Washington, Seattle, Washington 98195 USA*
[2]*Department of Biological Science, Stanford University, Stanford, California 94305 USA*

Abstract. Very little is known about the structure of microbial communities, despite their abundance and importance to ecosystem processes. Recent work suggests that bacterial biodiversity might exhibit patterns similar to those of plants and animals. However, relative to our knowledge about the diversity of macro-organisms, we know little about patterns of relatedness in free-living bacterial communities, and relatively few studies have quantitatively examined community structure in a phylogenetic framework. Here we apply phylogenetic tools to bacterial diversity data to determine whether bacterial communities are phylogenetically structured. We find that bacterial communities tend to contain lower taxonomic diversity and are more likely to be phylogenetically clustered than expected by chance. Such phylogenetic clustering may indicate the importance of habitat filtering (where a group of closely related species shares a trait, or suite of traits, that allow them to persist in a given habitat) in the assembly of bacterial communities. Microbial communities are especially accessible for phylogenetic analysis and thus have the potential to figure prominently in the integration of evolutionary biology and community ecology.

Key words: bacteria; community structure; microbial ecology; phylogenetic clustering and overdispersion; phylogenetic diversity; phylogenetic structure; relatedness.

Introduction

Although there may be millions of species of bacteria on Earth, we are only beginning to investigate patterns in their diversity (Horner-Devine et al. 2004*a*). Understanding patterns of bacterial diversity is of particular importance, because bacteria likely comprise the majority of the planet's biodiversity, they mediate many environmental processes that sustain life, and their diversity is of great importance in medicine, agriculture, and industry. Recent evidence suggests that bacteria can exhibit patterns in taxonomic diversity and community composition similar to those of plants and animals (e.g., Horner-Devine et al. 2003, 2004*b*). However, most of these studies have relied on measures of diversity that do not consider phylogenetic relatedness (Bohannan and Hughes 2003), and few studies have quantitatively examined bacterial communities within a phylogenetic context (Martin 2002, Davelos et al. 2004, Martin et al. 2004).

Phylogenetic measures can reveal differences in the richness or composition of two communities that would be identical using standard measures of species richness and composition (Martin 2002). Phylogenetic analyses of diversity have proven valuable in studies of plant and animal diversity, because such an approach can lend insight into the relative importance of evolutionary and ecological forces in shaping communities (Elton 1946, Webb et al. 2002, Cavender-Bares and Wilczek 2003). The idea that closely related taxa are more likely to interact intensely with each other than with more distantly related taxa is an old one (Darwin 1859); more recently, this idea has been expanded to suggest that interspecific interactions are influenced by the net ecological similarity of taxa, and closely related taxa tend to be more similar ecologically than distantly related taxa (Harvey and Pagel 1991). For example, co-occurring rainforest tree species have been observed to be more closely related than expected by chance (Webb 2000); such a pattern of phylogenetic attraction or clustering can indicate that these closely related taxa share traits important for their persistence in a particular environment (Webb et al. 2002). Such habitat filtering is important and might be more important than competition, in maintaining rain forest tree species diversity (see also Tofts and Silvertown 2000, Webb 2000, Kembel and Hubbell 2006).

In contrast, a community could be composed of distantly related taxa as a result of current or past competitive exclusion between similar (and thus closely related) taxa and/or as a result of convergent evolution in traits important for persistence in a given environment (Cavender-Bares et al. 2004, Kembel and Hubbell 2006). However, even for macro-organisms, relatively few studies have quantitatively examined community structure in a phylogenetic framework (but see other articles this issue), and even fewer have done so for

Manuscript received 9 February 2005; revised 13 July 2005; accepted 12 August 2005; final version received 7 September 2005. Corresponding Editor (ad hoc): J. B. Losos. For reprints of this Special Issue, see footnote 1, p. S1.
[3] E-mail: mchd@u.washington.edu

microbes (but see Martin 2002, Francis et al. 2003, Martin et al. 2004). We know of no other microbial studies that employ the approach used here to quantify community structure.

Bacteria offer a unique opportunity to examine the phylogenetic structure of multiple communities, because most recently published bacterial diversity data are molecular in nature and thus can be more easily interpreted within a phylogenetic context than data from many community studies of macro-organisms (Bohannan and Hughes 2003). A large proportion of microbes cannot be cultured with current laboratory techniques (Brock 1987), and thus bacterial taxa are often identified from the sequences of indicator genes extracted from environmental samples (Stackebrandt and Goebel 1994).

Here we take a quantitative approach to examining the phylogenetic structure of bacterial communities from a number of different environments. We ask whether bacterial communities exhibit phylogenetic structure (e.g., significant degrees of clustering or overdispersion of taxa across a phylogenic tree) and whether such patterns vary along environmental gradients.

Methods

Data

We used existing bacterial sequence data from four different environments and for three different genes. We selected data sets that were of high resolution (e.g., cloned and sequenced DNA sequences, rather than gradient gel or restriction fragment length data), that were from extensively sampled communities (relative to many studies of bacteria diversity), and that were replicated or spanned ecologically interesting environmental gradients in aquatic, soil, and sediment habitats.

The first data set consists of partial 16S rDNA sequences (the most common indicator gene used for bacterial diversity studies) sampled from freshwater mesocosms that span a productivity gradient (Horner-Devine et al. 2003). Each mesocosm consisted of a 2 m diameter polyethylene cattle tank with a screen cover, filled with well water. Each mesocosm was inoculated from the same-pooled sample collected from six ponds in southern Michigan that spanned a natural gradient in primary productivity. A gradient of primary productivity was established across the mesocosms by maintaining otherwise identical mesocosms with different input concentrations of nitrogen and phosphorus. At the end of a 4-mo growing season, one composite water column sample was used to generate a clone library from each of the five mesocosms. We selected ~100 clones from each of these libraries and sequenced 500 nucleotides from the 5′ terminal of each clone (GenBank accession numbers DQ064816–DQ065575).

The second data set consists of 16S rDNA sequences sampled from soil communities at different depths that differed in water saturation and total organic carbon (Zhou et al. 2002). Soil cores were collected from previously described sites in northern Virginia (Abbot's Pitt) and central Delaware (Zhou et al. 2002) at both the soil surface (depth = 0.05 m; GenBank accession numbers AY280351–AY289492) and subsurface (approximately at the depth of the water table, depth > 4.0 m; GenBank accession numbers AY456755–AY456883 and AY456885–AY456903) for a total of five samples (note that two samples were collected from the subsurface of Dover Air Force Base; hereafter, subsurface-D1 and subsurface-D2).

The third data set consists of 16S rDNA sequences sampled specifically from ammonia-oxidizing bacteria in Costa Rican soils (Carney et al. 2004). Samples were collected from three land use types on sandy loam soils: forest, pasture, and tree plantations. The tree plantation sites included three different site types that differed in plant community composition and richness. The one-species sites contained only *Cordia alliodora*, the three-species sites contained *C. alliodora*, an herb, and a palm, and the five-species sites contained *C. alliodora*, two palm species, and two other hardwoods. Each of the five site types (forest, pasture, and the three plantation treatments) was replicated three times, with the exception of the five-species site, which had two replicates. For each site type, a composite soil sample was collected from each of the replicate plots for a total of 14 samples. Partial 16S rDNA sequences from these samples were deposited in GenBank under accession numbers AY631475–AY631851.

The fourth data set consists of sequences of functional genes amplified from five sediments samples collected along a salinity and nitrogen gradient in the Chesapeake Bay (Francis et al. 2003; C. A. Francis, J. C. Cornwell, and B. B. Ward, *unpublished manuscript*). One of these functional genes (*amo*A) codes for a subunit of ammonia monooxygenase, an enzyme found only in ammonia-oxidizing bacteria (bacteria that mediate the transformation of ammonia into nitrite). A 450 base pair (bp) region was chosen for phylogenetic analysis, representing 150 amino acids (GenBank accession numbers AY352899–AY353054; Francis et al. 2003). The second functional gene (*nir*S) codes for a subunit of nitrite reductase, an enzyme found in denitrifying bacteria (bacteria that mediate the transformation of nitrite into nitrogen gas). A 233-bp region was used for analyses (C. A. Francis, J. C. Cornwell, and B. B. Ward, *unpublished manuscript*). Both gene fragments span the active site of their respective proteins (Berks et al. 1995, Rotthauwe et al. 1997, Braker et al. 2000).

For each data set, we screened sequences for chimeras and aligned them using the 2002 version of the ARB software package (for 16S genes; *available online*)[4] or Sequencher software (for functional genes; Gene Codes Corporation, Ann Arbor, Michigan, USA). We used only unambiguously aligned positions to construct the

[4] ⟨http://www.arb-home.de/⟩

phylogenetic hypotheses, and duplicate sequences were not used when generating phylogenetic trees. Thus sample sizes represent the number of unique sequences observed, rather than the total number of sequences analyzed.

Analysis

We used indices of community phylogenetic structure to compare these communities (Webb 2000). The net relatedness index (NRI) and nearest taxa index (NTI) measure the degree of phylogenetic clustering of taxa across a phylogenetic tree in a given sample relative to the regional pool of taxa. Positive values indicate that a community is clustered, whereas negative values indicate that community members are evenly spread or overdispersed across a phylogenetic tree. In other words, a positive NRI value indicates a community where members are on average more closely related to one another than they are to members of the regional taxon pool. Such a community thus appears to be clustered on a phylogenetic tree of the regional taxa. The NRI measures overall clustering across the phylogeny as the average distance between all pairs of taxa in a community. Specifically, $\text{NRI} = -(X_{\text{net}} - X(n))/\text{SD}(n)$, where X_{net} is the mean phylogenetic distance, measured as the mean pairwise branch lengths and thus a measure of pairwise sequence divergence, between all pairs of n taxa in a particular community; $X(n)$ and $\text{SD}(n)$ are the mean and standard deviation of phylogenetic distance for n taxa randomly distributed on the phylogeny. We obtain these latter values by 1000 random draws from the entire pool of taxa in the phylogeny. Alternatively, NTI measures the extent of terminal clustering on the phylogeny by determining the minimal distance or branch length between taxa in a particular community. The two indices are calculated similarly, except NTI substitutes X_{near} for X_{net}, where X_{near} is the shortest mean distance between all pairs of n taxa in a community sample. We calculated NRI and NTI using Phylocom 3.22 (Webb et al. 2005).

High and positive values of these indices indicate clustering of taxa across the overall phylogeny, whereas low or negative values indicate overdispersion of taxa across the phylogeny. We tested whether these values (and thus whether the extent of clustering) significantly differed from that of a randomly assembled community with a null model (1000 permutations of randomly drawn communities). We used a two-tailed significance test to evaluate the rank of observed values at $P = 0.05$, such that an observed rank of <25 or >975 was assumed to be significant overdispersion or clustering, respectively.

Calculation of NRI and NTI relies on a community phylogeny. We used ModelTest 3.06 to determine the best models of sequence evolution for the unique *amo*A and *nir*S sequences from the Chesapeake Bay sediment samples (Posada and Crandall 1998). Using the Akaike Information Criterion (Akaike 1973), we selected K81uf+I+G as the best model of sequence evolution for the *amo*A sequences and TVM+I+G for the *nir*S sequences. The 16S rDNA trees were constructed using neighbor-joining distance clustering with a HKY + gamma substitution model (Hasegawa et al. 1985), where gamma was estimated from the data. We used PAUP* to construct trees for all data sets (Swofford 2002). Maximum likelihood methods were used to estimate branch lengths based on the above HKY and gamma DNA substitution models. Trees were bootstrapped to examine phylogenetic robustness.

We also examined how phylogenetic clustering varies along environmental gradients. NRI and NTI values were standardized by the mean expected value for the number of taxa found in each community (Webb et al. 2002). We then used regression and ANOVA implemented in JMP, version 4.0, to examine the relationship between clustering values and environmental parameters (Sokal and Rohlf 1995). Where data were not normally distributed (as determined by the Shapiro-Wilk W test), we used the Kruskal-Wallis one-way analysis of variance by ranks (Sokal and Rohlf 1995). The following environmental parameters were considered: chlorophyll *a* (for the mesocosm data from Horner-Devine et al. [2003]), carbon content (for the soil community data from Zhou et al. [2002]), plant diversity and ammonia (for the ammonia oxidizer data from Carney et al. [2004]), ammonia and salinity (for the *amo*A data), and nitrate and salinity (for the *nir*S data). These environmental parameters were identified in the respective prior studies as important to taxonomic richness and/or community composition.

Results

Phylogenetic structure

We observed that most of the bacterial communities we examined exhibited significant phylogenetic structure (i.e., bacteria tended to co-occur with other bacteria that were more closely related than expected by chance). For example, bacterial communities from freshwater mesocosms exhibited significant and positive net relatedness index (NRI) and nearest taxa index (NTI) values (Table 1). This was true when all bacteria were considered, as well when the three most common groups of bacteria sampled from each of the mesocosms (Alphaproteobacteria, Betaproteobacteria, and Cytophaga-Bacteroides-Flavobacteria, or CFB) were considered separately. While there was some variation in relatedness among the different communities and groups of taxa, bacteria in the three most common taxonomic groups in these communities tended to be clustered.

Soil communities sampled at different depths showed different patterns of phylogenetic structure (Table 2). Subsurface soil communities showed significant clustering for both NRI and NTI. In contrast, one surface sample was randomly structured phylogenetically, and the other exhibited significant overdispersion for both indices.

TABLE 1. Net relatedness index (NRI) and nearest taxa index (NTI) results for the 16S rDNA sequences from the five freshwater mesocosm communities.

Community	*N*	NRI	NRI_gt	NTI	NTI_gt
All bacteria					
1	108	1.47*	934	5.54*	999
2	114	7.28*	999	4.95*	999
3	87	2.47*	988	2.81*	998
4	117	−2.18*	13	3.18*	998
5	104	3.8*	998	2.51*	998
Betaproteobacteria					
1	18	2.34*	991	2.71*	999
2	35	4.62*	999	0.87	814
3	2	1.02	76	−1.36	77
4	19	0.47	690	1.89†	972
5	7	0.31	627	0.98	825
Alphaproteobacteria					
1	18	−1.48	74	2.44*	998
2	28	2.96*	999	1.97*	994
3	14	1.53	935	1.35	913
4	33	−2†	25	1.94*	994
5	42	4.03*	999	2.35*	998
Cytophaga-Bacteroides-Flavobacteria (CFB)					
1	26	3.44*	999	4.57*	999
2	20	2.52*	996	3.29*	999
3	13	1.21	889	0.48	673
4	24	2.38*	996	2.68*	996
5	7	0.08	504	−0.78	234

Notes: *N* = no. taxa in a community. NRI_gt and NTI_gt represent the number of times the observed NRI and NTI values for a community, respectively, were greater than the value for randomly permuted communities.

* Communities that are significantly structured at the $P = 0.05$ level.

† Communities that are significantly structured at the $P = 0.10$ level.

Ammonia-oxidizer communities from Costa Rican soils exhibited the most variation in phylogenetic structure of all the data sets considered (Table 3). While communities from forest soils showed no significant phylogenetic structure, pasture communities tended to be overdispersed. Communities from the experimental plant treatments with one, three, or five plant species tended to be phylogenetically clustered overall with no clear pattern in the NTI.

TABLE 2. The NRI and NTI results for the 16S rDNA sequences from soil communities at different depths.

Community	*N*	NRI	NRI_gt	NTI	NTI_gt
Subsurface A	43	3.15*	999	3.72*	999
Subsurface D1	27	2.59*	994	2.97*	999
Subsurface D2	20	3.81*	999	3.81*	999
Surface D	66	−3.4*	0	−1.88†	27
Surface A	65	−0.98	159	−0.29	382

Notes: Labels D and A refer to samples collected at Dover Air Force Base (Delaware, USA) and Abbot's Pitt (Virginia, USA), respectively. As two subsurface samples were collected at Dover Air Force Base, they are denoted D1 and D2. Other abbreviations and symbols are as in Table 1.

* Communities that are significantly structured at the $P = 0.05$ level.

† Communities that are significantly structured at the $P = 0.10$ level.

TABLE 3. The NRI and NTI results for the Costa Rican soil nitrifiers.

Group	Community	*N*	NRI	NRI_gt	NTI	NTI_gt
Forest	F1	18	−0.58	264	0.77	751
Forest	F2	18	0.05	508	0.77	755
Forest	F3	18	−0.46	314	1.4	937
Pasture	PF	20	−2.06*	26	−0.16	394
Pasture	PR	20	−2.42*	9	−2.41*	10
Pasture	PS	19	−4.71*	0	−3.76*	0
One	One1	29	−1.99*	23	1.53	959
One	One2	33	2.73*	997	1.84*	986
One	One3	34	2.32*	996	1.26	882
Three	Three1	36	2.97*	998	1.93*	987
Three	Three2	35	1.19	887	−0.45	332
Three	Three3	31	2.43*	995	1.8*	983
Five	Five1	30	0.91	802	0.3	588
Five	Five2	27	2.5*	998	2.25*	999

Notes: "One," "Three," and "Five" indicate the three plantation treatments containing the respective number of different species. Community labels describe the group and sample number. Other abbreviations and symbols are as in Table 1.

* Communities that are significantly structured at the $P = 0.05$ level.

Finally, sediment bacterial communities sampled at five sites in the Chesapeake Bay were phylogenetically structured (i.e., clustered) and contained less genetic diversity than a randomly assembled community (Table 4). This was true for both the ammonia-oxidizing bacteria and denitrifying bacteria sampled. All but one sample showed significant overall phylogenetic structure (as estimated by NRI). Interestingly, one community of denitrifying bacteria exhibited significant overdispersion as measured by NRI. Phylogenetic clustering measured by NTI was more common for denitrifying bacteria than for ammonia-oxidizing bacteria. Only one of the five

TABLE 4. The NRI and NTI results for the five Chesapeake sediment communities.

Community	*N*	NRI	NRI_gt	NTI	NTI_gt
*amo*A					
CB1	24	3.58*	999	−0.64	275
CB2	18	5.66*	999	2.79*	999
CB3	22	6.17*	999	0.47	688
CT1	26	5.04*	999	0.88	789
CT2	14	1.95*	965	1.56	940
nirS					
CB1	79	−4.66*	0	−0.32	363
CB2	46	5.13*	999	3.56*	998
CB3	53	4.27*	999	3.45*	999
CT1	75	0.39	631	2.45*	990
CT2	87	3.09*	999	5.54*	999

Note: Sampling stations were located in the Choptank River (CT) as well as in the main channel of the Chesapeake Bay (CB). CT1 was located in the upper Choptank, while CT2 was located in the lower Choptank. Main channel stations were located in the north bay (CB1), mid-bay (CB2), and south bay (CB3). Other abbreviations and symbols are as in Table 1.

* Communities that are significantly structured at the $P = 0.05$ level.

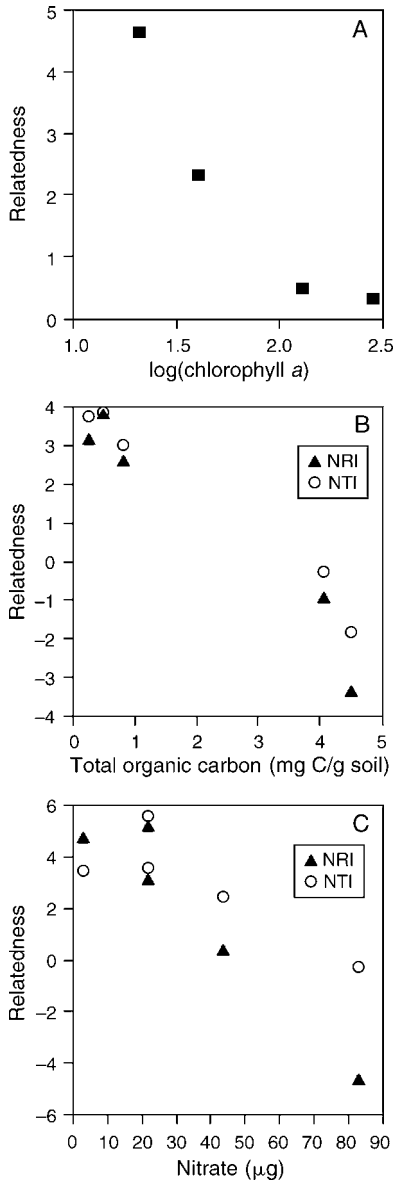

FIG. 1. Variation of relatedness along environmental gradients. (A) Relatedness decreased with increasing productivity in freshwater mesocosms for Betaproteobacteria (NRI, $R^2 = 0.829$, $P = 0.0588$; mesocosm three was excluded due to small community size). (B) Relatedness decreased with total organic carbon in bacterial soil communities (NRI, $R^2 = 0.925$, $P = 0.0058$; NTI, $R^2 = 0.966$, $P = 0.0017$). (C) There was a significant negative relationship between the relatedness of *nir*S genes and nitrate in Chesapeake Bay sediment communities (NRI, $R^2 = 0.909$, $P = 0.0077$; NTI, $R^2 = 0.601$, $P = 0.0768$).

ammonia-oxidizing communities sampled showed significant phylogenetic clustering as estimated by NTI.

Phylogenetic structure and the environment

We also observed that measures of phylogenetic structure can vary along environmental gradients. There was a negative trend between NRI estimated for Betaproteobacteria and primary productivity in aquatic mesocosms (Fig. 1A). Relatedness did not vary with productivity for any of the other bacterial groups examined in these mesocosms (results not shown).

Phylogenetic structure varied significantly with depth for soil communities (NRI, $t = 5.319$, $df = 3$, $P = 0.013$; NTI, $t = 6.693$, $df = 3$, $P = 0.0068$). In addition, communities sampled from soils with high total organic carbon had lower relatedness values than those sampled from low total organic carbon soils (Fig. 1B).

In the Chesapeake Bay, relatedness of denitrifying bacteria exhibited a nonsignificant trend with nitrate and salinity (Fig. 1C for nitrate results; salinity: NRI, NS, results not shown; NTI, $R^2 = 0.5121$, $P = 0.107$). In contrast, relatedness measures of ammonia-oxidizing bacteria did not vary with any environmental parameters measured (ammonium and salinity, results not shown).

Phylogenetic structure of communities of ammonia-oxidizing bacteria varied with plant community composition in Costa Rican soils for NRI (ANOVA, $F_{4,9} = 5.507$, MSE $= 2.33$, $P = 0.0160$), but not for NTI (Kruskal-Wallis: $\chi^2 = 6.5095$, $df = 4$, $P = 0.164$). Pairwise post-hoc comparisons of the different treatments revealed that the bacterial communities from sites with one, three, or five focal plant species were more clustered than pasture communities as measured by NRI. There was a weak positive relationship between ammonia and NTI, but not NRI, for these bacteria ($R^2 = 0.218$, $P = 0.052$).

DISCUSSION

Our results suggest that bacteria tend to co-occur with other closely related bacteria more often than expected by chance, as has been observed for some plant species (Webb 2000; also see Cavender-Bares et al. 2006, Kembel and Hubbell 2006, Lovette and Hochachka 2006, Weiblen et al. 2006). In addition, we observed that phylogenetic structure can vary along environmental gradients.

We observed significant net relatedness index (NRI) and nearest taxa index (NTI) values for freshwater bacterial communities from experimental mesocosms. Relatedness information provides a different window into bacterial communities than does information concerning richness or taxonomic composition. Accordingly, our observations that the relatedness of co-occurring Betaproteobacteria decreases with productivity, is in contrast to previous observations that their taxonomic richness does not (Horner-Devine et al. 2003). We observed that each of the five communities contained approximately the same number of taxa regardless of productivity, but these taxa tended to be more distantly related at higher productivities. Decreasing relatedness with increasing productivity might indicate that low productivity environments are more "stressful" (e.g., impose a stronger "filter" on a

community) for Betaproteobacteria than do more productive environments. In contrast, we did not observe a relationship between relatedness of Alphaproteobacteria or Cytophaga-Bacteroides-Flavobacteria (CFB) and productivity in the mesocosm study, despite our previous observations of changes in taxonomic richness of these groups with productivity (Horner-Devine et al. 2003).

Clustered distributions have been interpreted as evidence of habitat filtering, where a group of closely related species shares a trait, or suite of traits, that allow them to persist in a given habitat (Webb et al. 2002). Alternatively, significant phylogenetic clustering could be the result of differential dispersal and/or colonization abilities, or an adaptive radiation event. While it is beyond the scope of this work to determine the process responsible for the clustering we observed, the results from the freshwater mesocosms suggest that habitat filtering, rather than an adaptive radiation event or colonization effects, was important in the assembly of these bacterial communities. In this mesocosm study, bacterial communities were established in each mesocosm from the same inoculum of bacteria from natural pond communities (Horner-Devine et al. 2003). Thus, the history of colonization could not play a role in the patterns we observed. Although dispersal among mesocosms was not prevented (and likely occurred), the productivity gradient was randomized in space (i.e., mesocosms with similar productivities were not clustered in space), and thus clinal dispersal is unlikely to underlie the patterns we observed. Finally, since the mesocosm communities were sampled four months after they were initiated, it is unlikely that the ribosomal gene evolved during the course of the experiment, and thus it is unlikely that the clustering is due to adaptive radiation during the experiment. Without such information for the other data sets we analyzed, it is difficult to distinguish among habitat filtering, adaptive radiation, or colonization processes in the systems that were not manipulated. However, the results from the mesocosm study suggest that habitat filtering could be an important force in the assembly of at least some bacterial communities.

Overdispersion of taxa across a phylogeny has been observed in natural communities (Ackerly et al. 2006, Cavender-Bares et al. 2006, Silvertown et al. 2006) and could indicate that negative interactions (e.g., competition) are important in community assembly (Graves and Gotelli 1993, Webb et al. 2002). Although we did observe significant phylogenetic overdispersion in one of the freshwater bacterial communities, our observations do not suggest that competition played an overwhelming role in structuring the communities we studied at the scales examined here. However, it is important to note that habitat filtering and competition likely act in concert to produce the communities we observe. Thus even where the phylogenetic signature suggests the importance of habitat filtering, local competition can also be occurring.

Phylogenetic scale (the taxonomic group or rank under consideration) has been shown to influence the observation of phylogenetic patterns (Silvertown et al. 2001, 2006, Cavender-Bares et al. 2004, 2006). However, the effect of phylogenetic scale was not evident for bacteria from the freshwater mesocosms. We tended to observe phylogenetic clustering both when we examined all bacteria together and when we examined taxonomic subsets of the bacteria (Alphaproteobacteria, Betaproteobacteria, and CBF). It is possible that Alphaproteobacteria, Betaproteobacteria, and CBF encompass such a broad range of bacterial ecotypes that, even with the NTI (which focuses on terminal clustering and thus is particularly sensitive), it is less likely that one will observe overdispersion.

Recent work by Zhou et al. (2002) and Treves et al. (2003) suggests that spatial isolation plays an important role in structuring soil bacterial communities. They observed that unsaturated, surface communities had more uniform rank abundance patterns than did communities from saturated, subsurface communities, which exhibited high-dominance distributions (Zhou et al. 2002). They interpreted the uniform distribution from the surface samples as evidence that local competition does not play a significant role in structuring the soil communities they studied. Their observations (and subsequent mathematical modeling and laboratory experimentation) suggested that spatial isolation might limit competition in the surface soils (Treves et al. 2003). Our results do not support this hypothesis. We observed phylogenetic clustering in the subsurface samples, where spatial isolation was predicted to be minimized due to high water content and thus where there was an expectation for strong competition and phylogenetic overdispersion. In contrast we observed phylogenetic overdispersion in one of the surface samples, where isolation was predicted to be high and competition weak.

We did observe that phylogenetic clustering decreased with increasing total organic carbon (which covaried with depth) in the Zhou et al. (2002) soil data set. Thus as carbon availability increased, the strength of clustering and perhaps habitat filtering decreased. The potential decrease in the strength of filtering is unlikely to be related to an increase in the role of competition for carbon, since competition would likely decrease with increasing carbon. More information on the types of carbon present, as well as C:N ratios, might lend more insight into the underlying processes.

We observed both phylogenetic clustering and overdispersion for ammonia-oxidizing bacteria from Costa Rican soils. It is possible that, for ammonia-oxidizing bacteria, a more restricted (and potentially more ecologically similar) group of taxa, it might be easier to detect interactions among taxa, because ammonia-oxidizing bacteria likely compete for similar resources. Carney et al. (2004) found that neither bacterial richness nor composition changed across plant diversity treat-

ments (one, three, or five focal plant species). Similarly, we did not observe pairwise differences in phylogenetic structure among the plant treatments using post-hoc comparisons. Carney et al. (2004) also observed differences in the ammonia-oxidizer community among land-use types in some measures of diversity and in composition. We observed that, while pasture and forest did not differ in phylogenetic clustering, pasture communities were less phylogenetically diverse (i.e., less clustered) than each of the plant treatments. In fact, pasture communities tended to be overdispersed. We also found a weak positive relationship between terminal clustering and ammonia, such that clustering (and perhaps the importance of habitat filtering rather than competition) increased with ammonia. This is consistent with an increase in the importance of habitat filtering (and conversely, perhaps a decrease in the relative strength of competition) for ammonia as ammonia concentrations increased.

Analysis of *amo*A and *nir*S genes sampled from the Chesapeake Bay offered us an opportunity to examine community patterns deduced from potentially "ecologically relevant" genes. We assumed that phylogenetic overdispersion would be more prevalent in such data, where the taxa sampled are essentially from a single guild (i.e., a group performing the same function and requiring the same resources). However, we did not observe overdispersion for *amo*A, and we observed overdispersion in only one sample of *nir*S. In hindsight this is not overly surprising for *nir*S, given that denitrifying bacteria can be ecologically very different despite sharing the *nir*S gene (Shapleigh 2000). However, ammonia-oxidizing bacteria are believed to be a physiologically constrained group, existing solely on the oxidation of ammonia. The lack of overdispersion suggests that the *amo*A gene may be too conserved to reveal ecological differences or that sequence variation at this scale does not reflect ecological differences.

Previous analysis of the *amo*A samples demonstrated a nonsignificant trend toward decreasing richness with increasing salinity (Francis et al. 2003). We did not observe a relationship between phylogenetic structure as inferred from this gene and salinity or ammonia. In contrast, we observed a significant decrease in phylogenetic clustering as inferred from the *nir*S gene with increasing nitrate and a weak relationship between clustering and salinity. It is difficult to interpret the relationship with nitrate; one might expect that increasing nitrate availability would lead to decreased competition for this substrate by denitrifiers or stronger habitat filtering for denitrifying bacteria with genes that confer an advantage at high nitrate concentrations; we observed the opposite trend.

The interpretations of our results are based on the assumption that closely related taxa are more ecologically similar than distantly related taxa. This assumption has been shown to be true for some plants, animals, and microbes (Kuittinen et al. 1997, Nubel et al. 1999, Morgan et al. 2001, Prinzing et al. 2001), but not all (Losos et al. 2003, Rice et al. 2003, Knouft et al. 2006). How universally the assumption about similarity of closely related organisms applies to microorganisms is currently unknown. If this assumption does not hold for most bacteria, other explanations might be necessary for the patterns we observe. For example, some bacteria are capable of lateral gene transfer (LGT; Ochman et al. 2000, Lerat et al. 2003). Lateral gene transfer among co-occurring bacteria could weaken or uncouple the relationship between ecological similarity and evolutionary relatedness, if ecologically relevant genes are exchanged more often than phylogenetically informative housekeeping genes (e.g., ribosomal genes) as has been suggested (Lerat et al. 2003). Rampant LGT would reduce the prevalence of phylogenetic clustering or overdispersion due to ecological processes. However, we observed a significant level of phylogenetic clustering in the communities that we examined, suggesting that LGT does not substantially overwhelm phylogenetic patterns in these communities. In addition, recent work in environmental genomics suggests that on recent evolutionary time scales horizontal gene transfer is not rampant in natural microbial communities (Lerat et al. 2003).

Martin et al. (2004) used a different approach (lineage-per-time analysis) to look for phylogenetic patterns in microbial diversity data. They failed to show significant phylogenetic structure (i.e., an overabundance of closely related or distantly related sequences) across several different data sets. However their study differed from ours in that they assumed a "universal" null model (an exponential increase in lineages) for all data sets, rather than creating a null expectation for each data set by resampling of a regional phylogenetic tree. In the approach used here, we are interested in whether observed communities differ in phylogenetic diversity from communities created by a random draw from the available taxa in the regional pool. The Martin et al. (2004) approach suffers from a lack of power if the fraction of diversity sampled is small, making the task of detecting phylogenetic structure very difficult. While the communities examined here are also undersampled, the use of a relative measure of community relatedness decreases the influence of undersampling, provided communities in a given analysis are sampled with equal effort. Furthermore, if the questions of interest concern community assembly (as they do in our study), assuming a null model based on regional tree resampling is appropriate because it should model the assembly process.

We have observed that bacterial communities exhibit phylogenetic structure, in some cases similar to that observed for plants, and that this structure can vary along environmental gradients. Our results suggest that habitat filtering might be relatively more important to the assembly of bacterial communities than competition. Why might this be the case? Recent work suggests that

the greater the degree of environmental heterogeneity over which one samples a community, the more likely that phylogenetic clustering, rather than overdispersion, will be present (Cavender-Bares et al. 2006, Silvertown et al. 2006). The data sets used in our study consisted of samples that were extremely large relative to the size of the target organisms and the scales over which individuals interact (as is the case in most studies of microbial diversity) and thus likely included substantial environmental heterogeneity. As such, the spatial scale of sampling might bias the results towards clustering rather than overdispersion.

Recent studies have also suggested that phylogenetic scale should affect the prevalence of clustering; increasing phylogenetic scale (i.e., an increase in taxonomic lumping) might result in an increased prevalence of clustering (Silvertown et al. 2006). Substantial ecological diversity has been shown to be present within microbial taxa defined using molecular markers, especially ribosomal markers (e.g., Ward et al. 1998, Rocap et al. 2002). The molecular approach used to characterize microbial diversity might bias microbial diversity data sets toward detection of phylogenetic clustering when relatively large groups of organisms are targeted. Finer scale markers (e.g., internal transcribed spacer (ITS) or multi-gene approaches) might reveal the presence of increased phylogenetic overdispersion. The growing possibility of using environmental genomics to examine full genomes of different lineages from environmental samples will provide even more power to this approach (R. Whitaker and J. F. Banfield, *unpublished manuscript*).

The search for patterns in microbial biodiversity is in its infancy, and it is premature to make strong conclusions regarding the exact mechanisms responsible for the patterns we have described. To make such conclusions with confidence, a better understanding of the relationship between community assembly mechanisms and phylogenetic patterns is necessary. Such an understanding could be developed through studies of controlled experimental systems, where, for example, one can manipulate environmental parameters such as resource availability and tease apart the effects of mechanisms that occur on large scales of time and space (such as evolution and differential colonization) from those that occur on smaller scales (such as habitat filtering and competition). Microbial model systems could serve as excellent experimental systems in which to explore these ideas (Jessup et al. 2004). We also suggest that future work in natural systems examine gradients that span a greater range of environmental characteristics to test the hypothesis that phylogenetic clustering and thus habitat filtering increases with environmental extremes. Such studies coupled with a better understanding of the extent of LGT, how traits map onto phylogenies, and the evolutionary history of these traits (Ackerly et al. 2006, Silvertown et al. 2006) will help to explain the patterns of phylogenetic structure we observed. Where we do observe similar patterns of phylogenetic structure for microbial and macrobial communities, it is possible that similar mechanisms could be responsible for the structure of communities from these very different forms of life.

Acknowledgments

We are grateful to David Ackerly, Will Cornwell, Clara Davis, and Cam Webb for many informative discussions about phylogenetic structure and community assembly. We would also like to thank Karen Carney, Chris Francis, and Jizhong Zhou for allowing us to use their data for these analyses. Finally, we thank Jonathan Losos and two anonymous reviewers for their comments and suggestions. M. C. Horner-Devine was supported by funding from the American Association of University Women and the Center for Evolutionary Studies at Stanford. This work was also supported by a grant from the NSF (DEB0108556) to B. Bohannan.

Literature Cited

Ackerly, D. D., D. W. Schwilk, and C. O. Webb. 2006. Niche evolution and adaptive radiation: testing the order of trait divergence. Ecology **87**:S50–S61.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Academiai Kiado, Budapest, Hungary.

Berks, B. C., J. W. B. Moir, and D. J. Richardson. 1995. Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. Biochimica et Biophysica Acta **1232**:97–173.

Bohannan, B. J. M., and J. B. Hughes. 2003. New approaches to analyzing microbial biodiversity data. Current Opinion in Microbiology **6**:282–287.

Braker, G., J. Zhou, L. Wu, A. H. Devol, and J. M. Tiedje. 2000. Nitrite reductase genes (*nir*K and *nir*S) as functional markers to investigate diversity of denitrifying bacteria in Pacific northwest marine sediment communities. Applied and Environmental Microbiology **66**:2096–2104.

Brock, T. D. 1987. The study of microorganisms in situ: progress and problems. Symposium of the Society for General Microbiology **41**:1–17.

Carney, K. M., P. A. Matson, and B. J. M. Bohannan. 2004. Diversity and composition of tropical soil nitrifiers across a plant diversity gradient and among land-use types. Ecology Letters **7**:684–694.

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. American Naturalist **163**:823–843.

Cavender-Bares, J., A. Keen, and B. Miles. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. Ecology **87**:S109–S122.

Cavender-Bares, J., and A. Wilczek. 2003. Integrating micro- and macroevolutionary processes in community ecology. Ecology **84**:592–597.

Darwin, C. 1859. On the origin of species. John Murray, London, UK.

Davelos, A. L., K. Xiao, D. A. Samac, A. P. Martin, and L. L. Kinkel. 2004. Spatial variation in *Streptomyces* genetic composition and diversity in a prairie soil. Microbial Ecology **48**:601–612.

Elton, C. S. 1946. Competition and the structure of ecological communities. Journal of Animal Ecology **15**:54–68.

Francis, C. A., G. D. O-Mullan, and B. B. Ward. 2003. Diversity of ammonia monooxygenase (*amo*A) genes across environmental gradients in Chesapeake Bay sediments. Geobiology **1**:129–140.

Graves, G. R., and N. J. Gotelli. 1993. Assembly of avian mixed-species flocks in Amazonia. Proceedings of the National Academy of Sciences (USA) **90**:1388–1391.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford, UK.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution **22**:160–174.

Horner-Devine, M. C., K. M. Carney, and B. J. M. Bohannan. 2004*a*. An ecological perspective on bacterial biodiversity. Proceedings of the Royal Society of London B **271**:113–122.

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

Horner-Devine, M. C., M. A. Leibold, V. H. Smith, and B. J. M. Bohannan. 2003. Bacterial diversity patterns along a gradient of primary productivity. Ecology Letters **6**:613–622.

Jessup, C. M., R. Kassen, S. E. Forde, B. Kerr, A. Buckling, P. B. Rainey, and B. J. M. Bohannan. 2004. Big questions, small worlds: microbial model systems in ecology. Trends in Ecology and Evolution **19**:189–197.

Kembel, S. W., and S. P. Hubbell. 2006. The phylogenetic structure of a Neotropical forest tree community. Ecology **87**: S86–S99.

Knouft, J. H., J. B. Losos, R. E. Glor, and J. J. Kolbe. 2006. Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. Ecology **87**:S29–S38.

Kuittinen, H., A. Mattila, and O. Savolainen. 1997. Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. Heredity **79**:144–152.

Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-Proteobacteria. Public Library of Science **1**:101–108.

Losos, J. B., M. Leal, R. E. Glor, K. de Queiroz, P. E. Hertz, L. R. Schettino, A. C. Lara, T. R. Jackman, and A. Larson. 2003. Niche lability in the evolution of a Caribbean lizard community. Nature **424**:542–545.

Lovette, I. J., and W. M. Hochachka. 2006. Simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. Ecology **87**:S14–S28.

Martin, A. P. 2002. Phylogenetic approaches for describing and comparing diversity of microbial communities. Applied and Environmental Microbiology **68**:3673–3682.

Martin, A. P., E. K. Costello, A. F. Meyer, D. R. Nemergut, and S. K. Schmidt. 2004. The rate and pattern of cladogenesis in microbes. Evolution **58**:946–955.

Morgan, K. K., J. Hicks, K. Spitze, L. Latta, M. E. Pfrender, C. S. Weaver, M. Ottone, and M. Lynch. 2001. Patterns of genetic architecture for life-history traits and molecular markers in a subdivided species. Evolution **55**:1753–1761.

Nubel, U., F. Garcia-Pichel, M. Kuhl, and G. Muyzer. 1999. Quantifying microbial diversity: morphotypes, 16S rRNA genes, and carotenoids of oxygenic phototrophs in microbial mats. Applied and Environmental Microbiology **65**:422–430.

Ochman, H., J. G. Lawrence, and E. A. Groisman. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature **405**:299–304.

Posada, D., and K. A. Crandall. 1998. ModelTest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Prinzing, A., W. Durka, S. Klotz, and R. Brandl. 2001. The niche of higher plants: evidence for phylogenetic conservatism. Proceedings of the Royal Society of London B **268**: 2383–2389.

Rice, N. H., E. Martinez-Meyer, and A. T. Peterson. 2003. Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. Biological Journal of the Linnean Society **80**:369–383.

Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisholm. 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S–23S ribosomal DNA internal transcribed spacer sequences. Applied and Environmental Microbiology **68**:1180–1191.

Rotthauwe, J. H., K. P. Witzel, and W. Liesack. 1997. The ammonia monooxygenase structural gene *amo*A as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Applied and Environmental Microbiology **63**:4704–4712.

Shapleigh, J. P. 2000. The denitrifiers. *In* M. Dworkin, editor. The prokaryotes: an evolving electronic resource for the microbiological community. Springer-Verlag, New York, New York, USA.

Silvertown, J., M. Dodd, and D. Gowing. 2001. Phylogeny and the niche structure of meadow plant communities. Journal of Ecology **89**:428–435.

Silvertown, J., M. Dodd, D. Gowing, C. Lawson, and K. McConway. 2006. Phylogeny and the hierarchical organization of plant diversity. Ecology **87**:S39–S49.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research. Third edition. W. H. Freeman and Company, New York, New York, USA.

Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: a place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology **44**:846–849.

Swofford, D. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Massachusetts, USA.

Tofts, R., and J. Silvertown. 2000. A phylogenetic approach to community assembly from a local species pool. Proceedings of the Royal Society of London B **267**:363–369.

Treves, D. S., B. Xia, J. Zhou, and J. M. Tiedje. 2003. A two-species test of the hypothesis that spatial isolation influences microbial diversity in soil. Microbial Ecology **45**:20–28.

Ward, D. M., M. J. Ferris, S. C. Nold, and M. M. Bateson. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiology and Molecular Biology Reviews **62**:1353–1370.

Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. American Naturalist **156**:145–155.

Webb, C. O., D. D. Ackerly, and S. W. Kembel. 2004. Phylocom: software for the analysis of community phylogenetic structure and character evolution. Version 3.22. ⟨http://www.phylodiversity.net/phylocom⟩

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Annual Review of Ecology and Systematics **33**:475–505.

Weiblen, G. D., C. O. Webb, V. Novotny, Y. Basset, and S. E. Miller. 2006. Phylogenetic dispersion of host use in a tropical insect herbivore community. Ecology **87**:S62–S75.

Zhou, J., B. Xia, D. S. Treves, L. Y. Wu, T. L. Marsh, R. V. O'Neill, A. V. Palumbo, and J. M. Tiedje. 2002. Spatial and resource factors influencing high microbial diversity in soil. Applied and Environmental Microbiology **68**:326–334.