Ecology, 88(6), 2007, pp. 1345–1353

A COMPARISON OF TAXON CO-OCCURRENCE PATTERNS FOR MACRO- AND MICROORGANISMS

M. Claire Horner-Devine,[1,13] Jessica M. Silver,[1] Mathew A. Leibold,[2] Brendan J. M. Bohannan,[3] Robert K. Colwell,[4] Jed A. Fuhrman,[5] Jessica L. Green,[6] Cheryl R. Kuske,[7] Jennifer B. H. Martiny,[8] Gerard Muyzer,[9] Lise Øvreås,[10] Anna-Louise Reysenbach,[11] and Val H. Smith[12]

[1]*School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington 98195 USA*
[2]*Section of Integrative Biology, University of Texas, Austin, Texas 78712 USA*
[3]*Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403 USA*
[4]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269 USA*
[5]*Department of Biological Sciences, University of Southern California, California 90089 USA*
[6]*School of Natural Sciences, University of California, Merced, California 95344 USA*
[7]*Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 USA*
[8]*Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912 USA*
[9]*Department of Biotechnology, Delft University of Technology, 2628 BC Delft, The Netherlands*
[10]*Department of Biology, University of Bergen, N-5020 Bergen, Norway*
[11]*Department of Biology, Portland State University, Portland, Oregon 97201 USA*
[12]*Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045 USA*

Abstract. We examine co-occurrence patterns of microorganisms to evaluate community assembly "rules." We use methods previously applied to macroorganisms, both to evaluate their applicability to microorganisms and to allow comparison of co-occurrence patterns observed in microorganisms to those found in macroorganisms. We use a null model analysis of 124 incidence matrices from microbial communities, including bacteria, archaea, fungi, and algae, and we compare these results to previously published findings from a meta-analysis of almost 100 macroorganism data sets. We show that assemblages of microorganisms demonstrate nonrandom patterns of co-occurrence that are broadly similar to those found in assemblages of macroorganisms. These results suggest that some taxon co-occurrence patterns may be general characteristics of communities of organisms from all domains of life. We also find that co-occurrence in microbial communities does not vary among taxonomic groups or habitat types. However, we find that the degree of co-occurrence does vary among studies that use different methods to survey microbial communities. Finally, we discuss the potential effects of the undersampling of microbial communities on our results, as well as processes that may contribute to nonrandom patterns of co-occurrence in both macrobial and microbial communities such as competition, habitat filtering, historical effects, and neutral processes.

Key words: assembly rules; community composition; co-occurrence; microbial diversity; microbial ecology.

Introduction

Understanding the structure of organism assemblages is a central goal of ecology. Detecting patterns in assemblage structure over space and time can provide insight into the processes driving this structure and provides a context for experimental study of causal mechanisms. Here, we focus on one aspect of structure in community assemblages, taxon co-occurrence, and ask whether such patterns are similar for micro- and macroorganisms.

Longstanding interest among ecologists in co-occurrence patterns began to crystallize with the suggestion by Diamond (1975) that these patterns could reveal community assembly "rules" (Gotelli and Graves 1996). Diamond evaluated the distribution of bird species on islands in the Bismarck Archipelago. He argued that competitive interactions between species would result in nonrandom patterns of species co-occurrence, such that some taxa occur together more often than expected by chance or some taxa occur together less often than expected by chance. Such patterns can be evaluated by quantifying distributions, in which some species never co-occur (checkerboard distributions) while others do co-occur (permissible combinations), and comparing them to expectations from null models that assume that such patterns arise by chance. A recent meta-analysis by Gotelli and McCabe (2002) demonstrated that nonrandom co-occurrence patterns are common in macroorganisms. Specifically, they observed that macroorganism taxa deviate from random expectations across a range of spatial scales and that the degree of community structure varied by

Manuscript received 23 February 2006; revised 6 September 2006; accepted 20 September 2006. Corresponding Editor: R. B. Jackson. For reprints of this Special Feature, see footnote 1, p. 1343.
[13] E-mail: mchd@u.washington.edu

taxonomic group. For instance, assemblages composed of homeotherms demonstrated more nonrandom structure than poikilotherm assemblages, suggesting that physiological constraints may play a role in the assemblage structure.

Nonrandom co-occurrence patterns may be due to competitive interactions (as suggested by Diamond 1975) or other mechanisms, such as non-overlapping habitats or niches, mutualistic or syntrophic relationships, and/or historical effects (Gotelli and McCabe 2002). It is also possible that neutral processes may result in nonrandom co-occurrence patterns (Bell 2005). Nonrandom co-occurrence patterns thus may indicate that deterministic processes are important in structuring communities but do not identify the causal mechanisms responsible for such patterning. Although co-occurrence patterns similar to those described for macroorganisms by Gotelli and McCabe (2002) have not been quantified for microbial communities in this manner, qualitative observations of co-occurrence patterns do exist. These can differ from those expected to be important in macroorganisms, suggesting that the underlying processes may differ as well. One example of processes that may differ, at least in frequency, between microorganisms and macroorganisms is cooperation. The intimacy of cooperation can vary widely in microorganisms; however, in some of the extreme examples, one organism completely requires another to accomplish its metabolic activity, for example in methanogenic fatty acid degradation (Dong et al. 1994) and anaerobic methane oxidation (Boetius et al. 2000, Orphan et al. 2001). Alternatively, microorganisms and macroorganisms are also subject to many similar ecological processes and may thus show similarity in co-occurrence patterns. For example, microbes are subject to many of the same factors hypothesized to influence coexistence of macroorganisms (e.g., food web interactions, including competition and predation, habitat specialization, and so on) and interact strongly with macrobial organisms as well. It is currently unclear how such interactions might influence co-occurrence patterns. While we do not explicitly examine these underlying processes, our analyses provide a starting point for exploring the factors that determine assemblage structure.

Microorganisms (bacteria, archaea, fungi, protists, unicellular algae) differ from macroorganisms in many biologically relevant ways, the most obvious being body size, unicellularity, modes of reproduction, and physiology. However, recent work has revealed similarities in ecological patterns of microorganisms and macroorganisms (Horner-Devine et al. 2003, 2004*a*, *b*, 2006, Green et al. 2004, Bell et al. 2005, Smith et al. 2005, van der Gast et al. 2005, Green and Bohannan 2006, Horner-Devine and Bohannan 2006, Martiny et al. 2006). For example, both micro- and macroorganisms can exhibit significant turnover of taxa (beta-diversity) across landscapes (Hillebrand et al. 2001, Green et al. 2004, Horner-Devine et al. 2004*b*, Fierer and Jackson 2006, Martiny et al. 2006), which can result in a positive taxa–area relationship. In addition, communities of both micro- and macroorganisms can exhibit significant patterns of phylogenetic structure, with closely related taxa co-occurring less often than expected by chance (Webb 2000, Horner-Devine and Bohannan 2006, Kembel and Hubbell 2006).

To explore co-occurrence patterns for microorganisms, we follow the approach of Gotelli and McCabe (2002), who evaluated the application of several key aspects of Diamond's (1975) assembly rules to a variety of macroorganisms. First, we evaluate nonrandom patterns of taxa co-occurrence using three metrics. We then compare the microbial patterns to those observed for macroorganisms. Finally, we examine the influence of taxonomic group, habitat, and sampling method on the degree of taxon co-occurrence and conclude by discussing potential underlying processes.

Methods

Incidence matrices

We analyzed 124 data sets from microorganisms (see Plate 1) and contrasted the results with a previous analysis of 90 data sets of macroorganisms by Gotelli and McCabe (2002). We followed the procedures of Gotelli and McCabe (2002) as closely as possible given the data sets we obtained. In summary, each data set was arranged in an incidence matrix (i.e., a presence–absence matrix in which each row represents a species or taxon and each column represents a sampling site). Within each incidence matrix the presence of a species, taxon, or operational taxonomic unit (OTU) at a specific site was denoted by a 1, and its absence was denoted by a 0. If abundance data were collected, they were converted to presence–absence data.

Macroorganism data sets

The macroorganism data sets encompassed a range of habitats, from small quadrats in old fields to large islands in oceanic archipelagoes. The sampled taxa included birds, mammals, plants, and aquatic invertebrates, among others. Ninety of the original data sets were available for re-analysis so that we could confirm the use of the same analytical approaches.

The following criteria were used by Gotelli and McCabe (2002) to select and arrange these data sets: (1) data sets only included extant taxa (at the time of census); (2) data sets encompassing aquatic and terrestrial taxa were partitioned into separate matrices; (3) only the most inclusive data set from overlapping studies was used (e.g., if two different sampling methods were used to census, the data set with more species was used); (4) data sets from artificial substrates were excluded; and (5) conspecific subspecies were lumped together as a single taxon.

Microorganism data sets

We used the same criteria to select microorganism data sets from original literature sources as well as

TABLE 1. The number and taxonomic breakdown of microbial data sets.

Taxon	No. data sets
Total macrobes	90
Microbes	
Algae	4
Archaea	11
Bacteria	86
Ciliates	2
Eukarya	1
Fungi	12
Metazoa	1
Phytoplankton	2
Protist	5
Total	124

Note: When data sets did not differentiate among taxa, different taxonomic groups remained in one matrix (e.g., metazoa and eukarya).

directly from authors. Each data set included at least four sampling sites. The spatial scale of the data sets ranged from a study in which individual samples were taken from a single deep-subsurface borehole, to one in which individual samples were collected across North and South America. Data from experimental studies and/or those using artificial substrates, including field or greenhouse plots, lab cultures, mesocosms, and slurries, were excluded. However, data from control plots of larger experimental studies were used.

We divided the microbial data broadly by phylogenetic group (e.g., bacteria, archaea, and fungi; Table 1) and habitat (e.g., freshwater, marine, hot spring, coral, soil, sediment, vegetation, and animal waste). Data sets from studies that encompassed multiple habitats or taxonomic groups were partitioned into separate matrices so that each matrix only represented one taxonomic group and one habitat type.

A variety of molecular techniques and molecular markers were used to assess microbial assemblage composition. The methods used PCR-amplification followed by denaturing gradient gel electrophoresis (DGGE), terminal restriction fragment length polymorphism (T-RFLP), and cloning and sequencing. The molecular markers included 16S ribosomal RNA and various functional genes. It was not necessary to determine specific taxon identities for each OTU in these analyses; OTUs within a given data set were consistently designated (e.g., all using the same percentage similarity cutoff for sequences). As with macroorganisms, we used the most inclusive data set available from each study site. For example, when considering data sets where multiple molecular techniques or markers were used, we chose the one that detected the most operational taxonomic units (OTUs).

Quantifying species co-occurrence

We used three indices to quantify patterns of taxa co-occurrence for each of the data sets independently. First, the checkerboard index (C-board) calculates the number of species pairs that never occur together (so called because of the checkerboard units formed in an incidence matrix when each of two taxa occurs without the other, in a particular pair of samples; Table 2). Each such species pair forms a basic checkerboard unit (CU), i.e., a 2×2 submatrix of the form 10/01 or 01/10. The C-board score is based on the number of perfect CUs in an incidence matrix.

A checkerboard distribution is the simplest pattern of co-occurrence that may suggest the role of competition or differential habitat utilization in structuring communities. In other words, we might interpret a CU in which a pair of taxa never co-occur as evidence for competitive exclusion. It is equally possible to interpret a CU as evidence of differences in habitat preferences between taxa (with or without competition). According to Diamond's (1975) assembly rules, a community that is organized based on interspecific competition should have more checkerboard units than expected by chance, and the same pattern could arise from habitat partitioning. However, because this measure of co-occurrence relies on strict checkerboards, it is less informative than other, less conservative indices as it does not include information on pairs of taxa that co-occur less often than expected by chance, but do not segregate into perfect CUs.

A second community structure index (Combo) is the number of unique species combinations found between pairs of sites (Gotelli and McCabe 2002). Diamond's assembly rules describe "permissible combinations" and "forbidden combinations," both of which can include one or more species pairs. In a competitively structured community there should be fewer unique species combinations than expected by chance. We expect the same for a community dominated by habitat filtering, where different taxa have different and non-overlapping habitat preferences or tolerances.

Finally, we used Stone and Robert's (1990) C-score, which is based on the average number of checkerboard units (CUs) between all possible species (or taxon) pairs in a matrix. The C-score thus measures species segregation, but it does not require perfect checkerboard distributions. Stone and Roberts (1990) showed that the number of CUs for any species pair can be calculated as: $\mathrm{CU} = (R_i - S)(R_j - S)$ where R_i and R_j are the row totals for species i and j, and S is the number of sites containing both species. The C-score is then calculated as the mean number of CUs per species pair, for all

TABLE 2. A 2×2 submatrix of a larger incidence matrix showing a checkerboard unit (CU), in which a pair of taxa do not co-occur.

	Checkerboard unit (CU) in an incidence matrix	
Taxon	Sample A	Sample B
1	1	0
2	0	1

TABLE 3. Habitats sampled for the three most common microbial taxa in the meta-analysis ($n = 109$).

Habitat	Archaea	Bacteria	Fungi
Freshwater	0	5	0
Marine	8	32	0
Other	1	5	2
Sediment	1	12	2
Soil	1	32	8
Total	11	86	12

species pairs in a data set. If the resulting C-score is significantly larger than the C-score produced by the null distribution then at least some pairs of species co-occur less often than expected by chance (i.e., there is segregation of taxa). The reciprocal is also true. If the C-score is significantly less than the C-score for the null distribution then more species co-occur than expected by chance (i.e., there is more aggregation among taxa). We expect that an assemblage dominated by competitive species interactions would exhibit a significant segregation of taxa, whereas one dominated by mutualistic or syntrophic interactions would exhibit an aggregation of taxa (i.e., have a small C-score). Again, it is possible that habitat filtering, neutral processes, or phylogenetic constraints could also produce these patterns.

The C-score index is related to, but less stringent than the C-board index, as it does not require perfect segregation or association between OTUs. In other words, the C-score index can detect nonrandom patterns in a real matrix that does not display a perfect checkerboard distribution for particular pairs of taxa.

Null model algorithm

We tested the observed values for each of the three indices for significance using a Monte Carlo "null model" to randomize each matrix in the data set. Following the most conservative option for null model comparisons, the row and column totals from the original matrix were retained, thus conserving the number of species per site and sites per species. A swap algorithm in which randomly selected 2×2 submatrices of the form 01/10 or 10/01 (CUs) were exchanged was used to generate random matrices (Gotelli and Entsminger 2005). A swap algorithm begins with the original matrix and creates new matrices by repeatedly swapping. For each original data set, 5000 initial swaps randomized the original pattern, yielding one random template matrix followed by 5000 consecutive swaps to create 5000 unique null matrices. The co-occurrence index was calculated for each simulated matrix (I_{sim}). We then tested whether the observed co-occurrence indices (I_{obs}) significantly differed from that of simulated assemblages, using a two-tailed significance test to evaluate the rank of observed values at $P = 0.05$, such that an observed rank of less than 250 (i.e., 5% of 5000, indicating significant association among taxa) or greater than 4750 (i.e., 95% of 5000, indicating significant segregation among taxa) was assumed to be significantly different from the null.

Meta-analysis (standardized effect size, SES)

Following Gotelli and McCabe (2002), we calculated a standardized effect size for each matrix that allowed us to compare the degree of co-occurrence across data sets. The standardized effect size measures the statistical amount of deviation from random co-occurrence. It is calculated as the number of standard deviations that the observed score is above or below the mean co-occurrence index for the simulated communities. In other words, we calculated an effect size by standardizing the difference between the observed incidence matrix (I_{obs}) and the mean of 5000 indices generated from the simulated null model matrices (I_{sim}) functions. I_{sim} represents the pattern expected from a community in the absence of species interactions, historical effects on co-occurrence, habitat segregation, or sampling effects. We used the formula $SES = (I_{obs} - I_{sim})/S_{sim}$, where S_{sim} is the standard deviation of 5000 simulated communities (Gotelli and McCabe 2002). We used a one-sample Wilcoxon signed rank test to determine if the median SES differed from zero. A large SES for C-score, C-board, or Combo is the equivalent of a large index value for each respective index.

We used the default setting of the EcoSim (7.72) software to conduct all null model analyses and to calculate the co-occurrence indices and corresponding SES for each matrix (Gotelli and Entsminger 2005).

Of the three indices, the C-score has been shown to be the most sensitive measure of nonrandomness and to have the greatest statistical power (Gotelli 2000). We thus examined how variation in the SES of the C-score was related to taxonomic group, habitat, and method used to assess community similarity. In particular, we examined the influence of habitat (Table 3) on community structure for archaea, bacteria, and fungi, the three best-represented taxa in the meta-analysis. Habitats were classified as aquatic (freshwater or marine), sediment/soil, or other (including rock, hot spring, and so on; see Appendix A). The SES distributions were not normally distributed for microbes or macrobes (results not shown); therefore the nonparametric Kruskal-Wallis test for the difference among medians was used to examine the effects of taxon and habitat on the SES C-score.

Because of the vast diversity of microbial taxa and the limited ability of many fingerprinting techniques to sample that diversity exhaustively, many of the microbial assemblages included in this analysis are likely to be undersampled relative to the macroorganism data sets, most of which were collected as a result of thorough census efforts. To explore the effect of potential undersampling on the microbial assemblage scores, in comparison with macrobial assemblages, we simulated undersampling of our macroorganism data sets and compared the C-scores to the microorganism data sets. We subsampled 12 macroorganism data sets by drawing

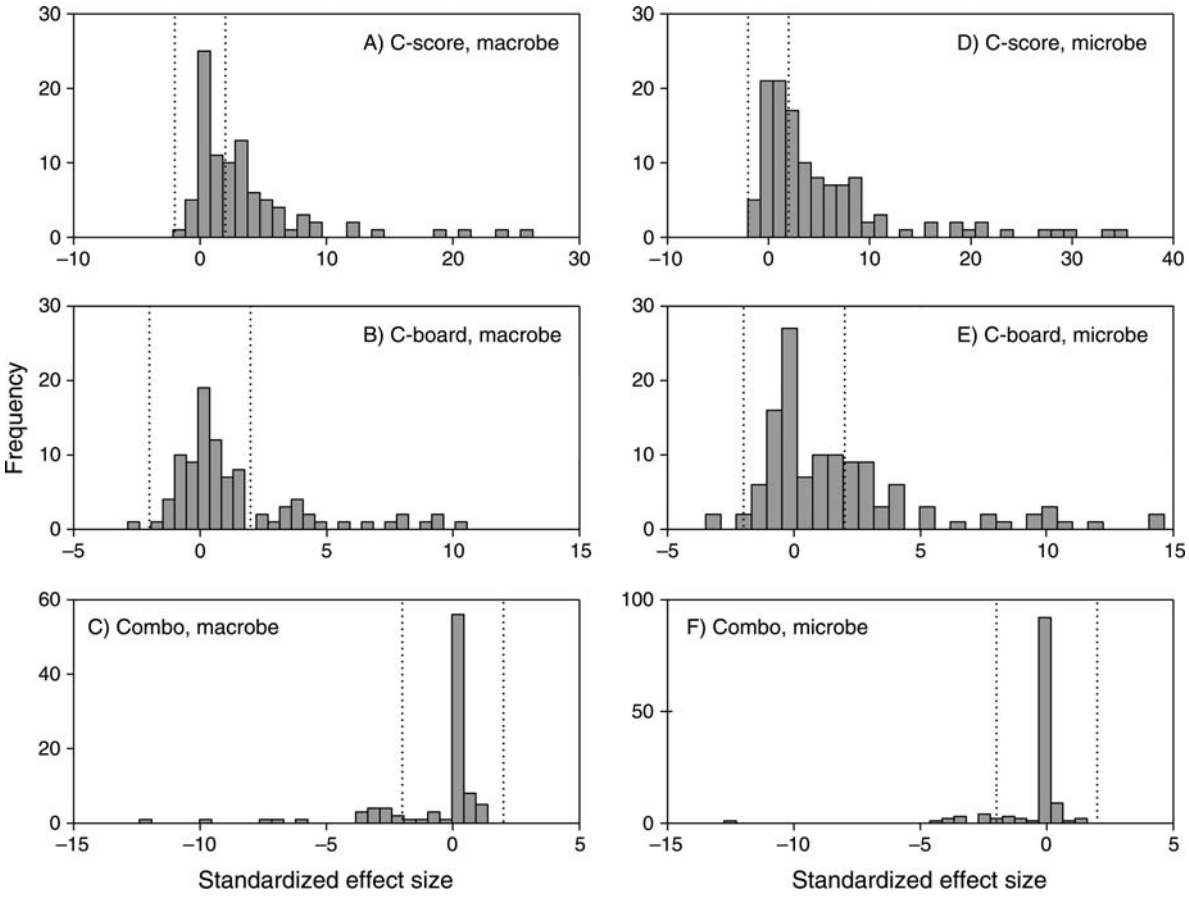

FIG. 1. Frequency distributions for standardized effect sizes for the three co-occurrence index measures (described in *Quantifying species co-occurrence*) from the macrobial ($n = 90$; left-hand panels) and microbial ($n = 124$, right-hand panels) incidence matrices. The null hypothesis is that the average effect size is zero and that 95% of the observations will lie between −2 and +2. (A, D) The standardized effect size (SES) for the C-score (a measure of species segregation) was significantly greater than expected for both macrobes (mean = 3.61, $H = 1811.5$, $P < 0.0001$) and microbes (mean = 5.35, $H = 3365.5$, $P < 0.0001$). (B, E) The C-board SES was significantly greater than expected, based on the null distribution for macrobes (mean = 1.37, $H = 912$, $P < 0.0001$) and for microbes (mean = 1.83, $H = 1800.5$, $P < 0.0001$). (C, F) The SES for the number of species combinations also was significantly greater than expected for macrobes (mean = −0.80, $H = -253$, $P < 0.008$) and for microbes (mean = −0.4006, $H = -181$, $P < 0.010$). Vertical dotted lines at +2 and −2 represent 95% confidence intervals for the null hypothesis that the average SES equals zero.

100 random subsets of taxa from each data set. For example, for a given observed macroorganism incidence matrix, we generated a new matrix that subsampled 20% of the original species with equal probability (without replacement) and then calculated the co-occurrence index. This new matrix thus simulated an undersampling of the macroorganism assemblages. We repeated this 100 times for each of 20%, 40%, 60%, and 80% of the original number of macrobial species observed in a given data set. We used EcoSim to calculate the C-score and C-score SES for each of these matrices and then plotted the average SES C-score for each matrix at each subsampling level.

RESULTS

Microbial assemblages displayed a substantial proportion of nonrandom co-occurrence patterns as measured by the C-score, with 56% of the 124 data sets showing significantly more segregated structure than expected by chance and only one data set showing significantly more aggregation than expected by chance. This is broadly similar to patterns for macroorganisms, which also show more nonrandom co-occurrence patterns structure than expected by chance (49% of 90 data sets).

The standardized effect score (SES) for the C-board index was significantly greater than zero for both microorganisms and macroorganisms, whereas the SES for the number of species combinations (Combo) was significantly less than zero for both groups (Fig. 1). Similarly the SES for the C-score differed significantly from the expected distribution for both macroorganisms and microorganisms. Thus overall both groups of organisms exhibited more assemblage structure than

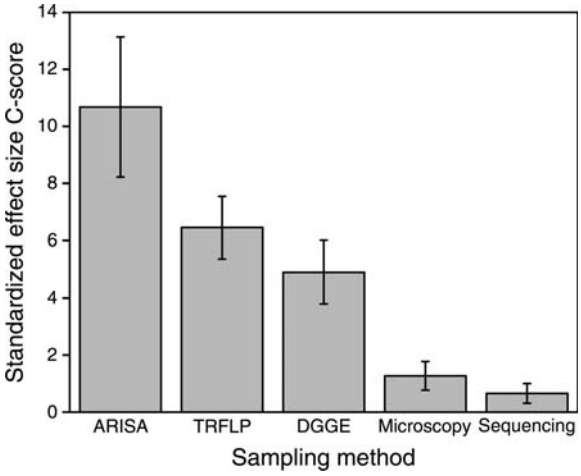

FIG. 2. The influence of molecular method on C-score standardized effect size, including all molecular markers. Automated ribosomal intergenic spacer analysis (ARISA) resulted in the highest values, whereas clone libraries and sequencing resulted in the lowest. Also included in the analysis were studies using terminal restriction fragment length polymorphism (T-RFLP) and denaturing gradient gel electrophoresis (DGGE).

expected by chance and showed broadly similar patterning.

Despite the many differences between the biology of microorganisms and macroorganisms, as well as how they are sampled, their mean C-score SES did not differ (Kruskal-Wallis, df = 1, $H = 1.77$, $P = 0.184$). The same was true for the mean C-board SES (Kruskal-Wallis, df = 1, $H = 0.889$, $P = 0.346$) and the Combo SES (Kruskal-Wallis, df = 1, $H = 0.0210$, $P = 0.885$). These results are especially interesting given the different ways taxa are defined and the often greater ecological breadth represented by such microbial taxa.

The C-score SES did not vary with the size of the data set, as measured by the number of samples or the number of different OTUs included in a given incidence matrix (results not shown). In contrast with the macroorganism data, the C-score SES also did not vary among bacteria, archaea, and fungi; the three best represented taxa in the analysis (Kruskal-Wallis, df = 2, $H = 0.3303$, $P = 0.85$). Because only limited data sets were available for archaea and fungi, we examined the influence of habitat on co-occurrence for bacteria only. The bacterial C-score SES did not vary with habitat for freshwater, marine, soil, and sediment habitats (Kruskal-Wallis, df = 3, $H = 1.96$, $P = 0.58$).

Because taxa and habitat did not influence the C-score SES, we pooled the microbial taxa and habitats and examined the effect of the type of molecular analysis (i.e., T-RFLP, cloning/sequencing, DGGE, and so on) as well as the molecular marker used (i.e., rRNA gene, functional gene, and so on). Sampling method had a significant effect on SES C-score for microbes (Kruskal-Wallis, df = 4, $H = 44.37$, $P < 0.0001$; Fig. 2). This result is not surprising as different molecular approaches sample assemblages at different taxonomic resolutions.

The effect of the molecular marker examined depended on the type of molecular analysis. When cloning and sequencing were used, the C-score SES for data sets targeting ITS, ribosomal, or functional genes did not differ (Kruskal-Wallis, df = 2, $H = 2.5399$, $P = 0.281$). However, the molecular marker did have a significant effect on C-score SES for assemblages characterized by T-RFLP (Kruskal-Wallis, df = 2, $H = 10.08$, $P = 0.0065$), with functional genes resulting in lower C-score SES than ITS or ribosomal genes. We controlled for the effect of molecular analysis and marker and examined the effect of habitat on bacteria sampled using T-RFLP of the 16S rDNA. Again, habitat did not influence the C-score SES (results not shown).

Undersampling tends to underestimate the degree of segregation among species. Of 12 subsampled macroorganism matrices, eight showed an increasing trend of C-score SES with the increasing percentage of the taxa sampled (Fig. 3). Thus undersampling makes it more difficult to accept the hypothesis of significant segregation. Our results that many microbial assemblages are significantly segregated are thus all the more striking.

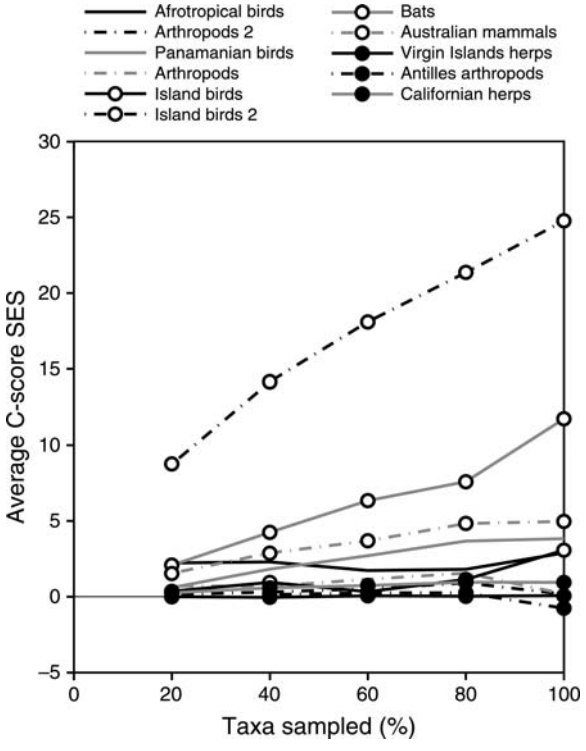

FIG. 3. The effect of undersampling on C-Score SES for macroorganisms. The SES tends to increase with the percentage of the taxa sampled. Different symbols correspond to a subset of different macrobial taxa from the macrobial data sets. Assuming that macrobial communities were exhaustively sampled, these patterns indicate the possible extent of bias due to undersampling that might occur in microbes.

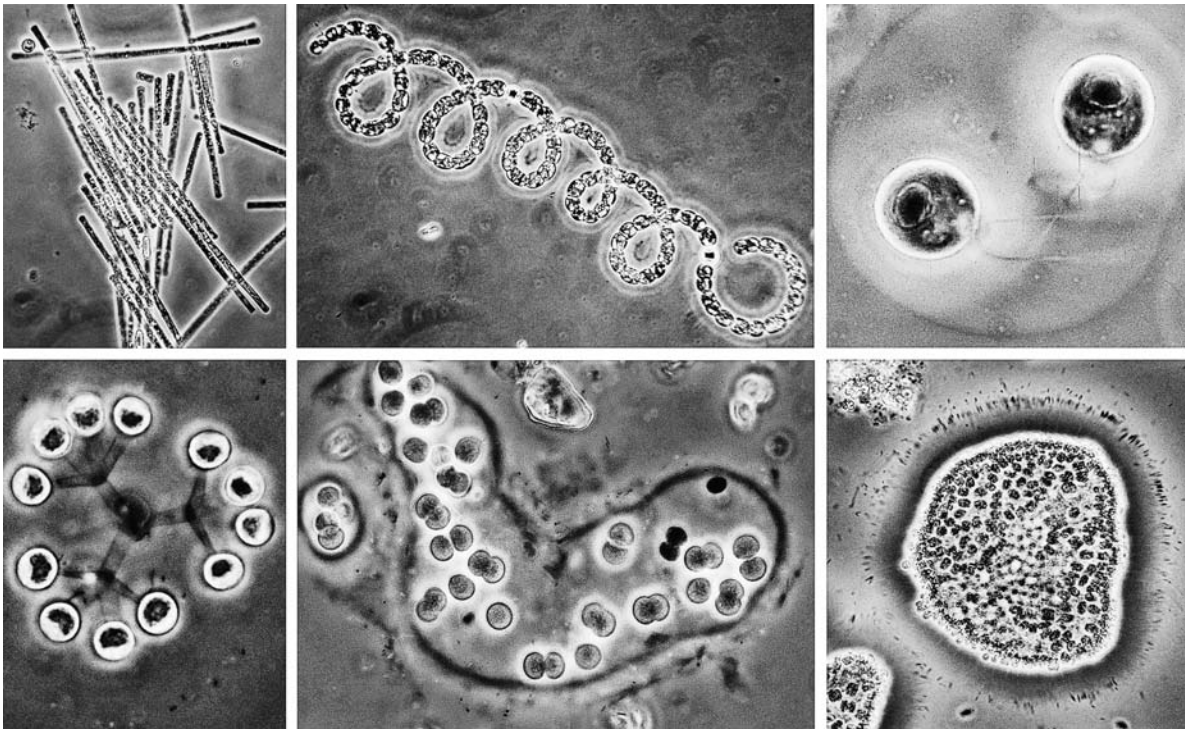

PLATE 1. We report on co-occurrence patterns of microorganisms in order to evaluate community assembly rules. The microbial plankton in aquatic ecosystems include diverse and beautiful taxa, such as (clockwise, from top left): *Aphanizomenon*, *Anabaena*, *Dictyosphaerium*, *Sphaerocystis*, *Microcystis*, and *Coelosphaerium*. Photo credit: R. O. Megard.

DISCUSSION

Co-occurrence patterns are interesting and informative ways of exploring assemblage structure. In macroorganisms these patterns commonly deviate from null expectations in ways that indicate how assemblages may be structured by interspecific competition and other mechanisms. Here we show that microbial assemblages, like those of macroorganisms, exhibit segregation (i.e., less co-occurrence) of taxa more often than expected by chance. We further show that metrics of these patterns show overall frequency distributions similar to those of macroorganisms, although undersampling of microorganisms may have underestimated structuring in microbial assemblages; approximately half of the microbial data sets analyzed showed significant co-occurrence structure, and very few displayed evidence for aggregation. When we compare our results with those observed for macroorganisms, we see that this general pattern of assemblage structure appears to be a common feature of communities in all three domains of life and is robust to differences in body size, spatial scale, and sampling methods. The degree to which similar patterns are due to similar causal mechanisms is impossible to evaluate without experimental work, but our findings suggest at least two possible explanations. First, they may indicate that similar processes structure both microbial and macrobial communities, as previously suggested (Horner-Devine et al. 2004*a*, Green and Bohannan 2006, Martiny et al. 2006). Alternatively, they may indicate that patterns of assemblage structure involving microorganisms and macroorganisms are causally interlinked such that the similarities reflect interdependencies rather than similar but independent mechanisms.

Diamond's (1975) predictions regarding co-occurrence patterns stemmed from the idea that interspecific competition would result in forbidden combinations of species and checkerboard distributions. Recent thinking suggests that other processes can contribute to such patterns whether or not competition is present. For instance, segregation patterns could result from "habitat checkerboards," a mechanism that may apply equally to micro- and macroorganisms. In other words, species or taxa may be associated with different environmental features of the sites, producing less co-occurrence than expected by chance. Other studies have demonstrated that microbial community composition can vary with environmental characteristics (reviewed in Horner-Devine et al. 2004*a*; see also Crump et al. 2007, Fierer et al. 2007). While we also cannot directly evaluate these mechanisms in this study, we believe that these phenomena likely influence the patterns we observe.

A third possible explanation is that historical or phylogenetic processes have led to more segregation of taxa than expected by chance. In particular, allopatric speciation may lead to a pattern of little or no coexistence among congeners, whether or not competi-

tion is occurring or occurred in the past between such species (Connell 1980). Recent work on both macrobial and microbial communities suggests that for a variety of communities (both microbial and macrobial) co-occurring taxa are less closely related than expected by chance (e.g., Cavender-Bares et al. 2006, Horner-Devine and Bohannan 2006, Weiblen et al. 2006). While the indices do not take phylogenetic distances into account, these previous results suggest that historical processes may play a role in structuring microbial assemblages.

These three explanations for nonrandom co-occurrence (competition, habitat checkerboards, and historical effects), and more specifically for segregation, are not mutually exclusive, and in fact, probably interact to reinforce these patterns. For example, competitive interactions may lead to the divergent evolution of distinct habitat preferences and a consequent reduction in niche overlap. Finally, recent work suggests that neutral processes may result in nonrandom co-occurrence patterns (Bell 2005).

Although our analysis did not detect differences in the degree of co-occurrence exhibited by communities of micro- and macroorganisms, it is possible that undersampling of the microbial communities masks differences that do, in fact, exist. Most microbial communities, especially in soil systems, are vastly undersampled. For example, Tringe et al. (2005) sequenced 1700 clones from clone libraries of PCR-amplified bacterial 16S rRNA sequences generated from a single sample of soil DNA. Richness estimators suggest that despite such intensive sampling, they were unable to sample more than 30% of the taxa present. Random subsampling of the relatively well-sampled macroorganism data sets (Fig. 3) suggests that undersampling may underestimate the degree of structure in a data set. Thus as most of the microbial data sets are likely to be undersampled, this result suggests that increased sampling effort might result in even higher C-scores for microbial assemblages and thus microbes might exhibit even more segregation than macroorganism assemblages, at the spatial scales of measurement represented in our data.

Neither taxonomic group nor habitat influenced the degree of co-occurrence for microbial taxa. In contrast, previous work demonstrated that the degree of co-occurrence differs significantly among taxonomic groups of macroorganisms (habitat was not examined; Gotelli and McCabe 2002). It is important to note, however, that the microbial groups examined were much broader ecological groups than were those considered for macroorganisms (bacteria, archaea, fungi vs. fish, ants, birds), and this discrepancy might also contribute to the differing results.

Whereas taxon identity and habitat do not appear to influence the degree of co-occurrence in microbial assemblages, the molecular method used for taxon identification did. This result suggests that taxonomic resolution might influence the co-occurrence index values. Collapsing macroorganism matrices from the species level to the genus level could represent an effect similar to sampling at different taxonomic levels and could be expected to reduce the level of apparent segregation, as checkerboard congeners are pooled into a single taxon for each genus. In addition, when assemblages represented in incidence matrices do not represent coherent ecological guilds and thus may include taxa that are not expected to show significant structure, the co-occurrence analyses here may be subject to a "dilution effect" (Diamond and Gilpin 1982). This is more likely to be the case with many of the microbial data sets than for the macrobial data sets. For example, assemblages queried using ribosomal markers likely represent a broad range of physiological traits. It is thus easy to imagine that this analysis has conservatively represented assemblage structure patterns. Future work could address this possibility by comparing data sets that differ only in the taxonomic breadth of markers used (i.e., one could compare data sets composed of beta-proteobacteria to data sets composed of beta-proteobacteria that are also ammonia-oxidizing bacteria). The method used to query the microbial populations did have an effect on co-occurrence scores. It is tempting to suggest that sampling method and resolution work in concert to influence co-occurrence scores, but such a pattern is not apparent. Sampling effort and the spatial scale of sampling may also interact to determine the values observed here, and this interaction deserves more study. It is also likely that taxonomic resolution and the degree of undersampling are confounded for the microbial data sets. In other words, we might expect that a technique that samples diversity at a relatively fine resolution will sample a higher degree of diversity. This might lead one to expect cloning and sequencing to yield relatively high co-occurrence values. We did not observe this here. It is possible that other factors such as the choice of molecular marker (here, often rRNA gene) and degree of undersampling can confound this expectation.

Our results suggest that assemblages composed of organisms from all domains of life exhibit more co-occurrence structure than expected by chance. We show that for microorganisms from a variety of taxonomic groups, sampled from a range of habitats, at different spatial scales, and using a broad array of sampling approaches, some taxa co-occur less often than expected by chance, just as has been shown for macroorganisms. In addition, when co-occurrence structure was detected (in over half the communities), this structure took the form of segregation, not aggregation (with one exception). For both micro- and macroorganisms, additional information on habitat preferences, phylogenetics, and sampling effort will be necessary to understand microbial assemblage structure and its underlying processes more fully.

ACKNOWLEDGMENTS

This work was conducted as part of the Patterns in Microbial Diversity Working Group supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded

by NSF (Grant No. DEB-00-72909), the University of California at Santa Barbara, and the State of California. We thank M. Kane, J.A. Krumins, P. Morin, S. Naeem, O. Petchey, J. Staley, and D. Ward for many useful and lively discussions, and J. Schimel and N. Fierer for comments on the manuscript. We especially thank the many research groups, including Microbial Observatories, who graciously contributed data to this analysis.

Literature Cited

Bell, G. 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86:1757–70.

Bell, T., D. Ager, J. Song, J. A. Newman, I. P. Thompson, A. K. Lilley, and C. J. van der Gast. 2005. Larger islands house more bacterial taxa. Science 308:1884–1884.

Boetius, A., K. Ravenschlag, C. J. Schubert, D. Rickert, F. Widdel, A. Gieske, R. Amann, B. B. Jorgensen, U. Witte, and O. Pfannkuche. 2000. A marine microbial consortium apparently mediating anaerobic oxidation of methane. Nature 407:623–626.

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

Connell, J. H. 1980. Diversity and the coevolution of competitors, or the ghost of competition past. Oikos 35: 131–138.

Crump, B. C., H. E. Adams, J. E. Hobbie, and G. W. Kling. 2007. Biogeography of bacterioplankton in lakes and streams of an Arctic tundra catchment. Ecology 88:1365–1378.

Diamond, J. M. 1975. Assembly of species communities. Pages 342–444 *in* M. L. Cody and J. M. Diamond, editors. Ecology and evolution of communities. Harvard University Press, Cambridge, Massachusetts, USA.

Diamond, J. M., and M. E. Gilpin. 1982. Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. Oecologia 52(1):64–74.

Dong, X., G. Cheng, and A. J. M. Stams. 1994. Butyrate oxidation by *Syntrophospora bryantii* in coculture with different methanogens and in pure culture with pentenoate as electron acceptor. Applied Microbiology and Biotechnology 42:647–652.

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

Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the National Academy of Sciences (USA) 103(3):626–631.

Gotelli, N. J. 2000. Null model analysis of species co-occurrence patterns. Ecology 81:2606–2621.

Gotelli, N. J., and G. L. Entsminger. 2005. EcoSim: null models software for ecology. Acquired Intelligence Inc. and Kesey-Bear, Jericho, Vermont, USA.

Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. Smithsonian Institution Press, Washington, D.C., USA.

Gotelli, N. J., and D. J. McCabe. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules. Ecology 83:2091–2096.

Green, J. L., and B. J. M. Bohannan. 2006. Spatial scaling of microbial biodiversity. Trends in Ecology and Evolution 21(9):501–507.

Green, J. L., A. J. Holmes, M. Westoby, I. Oliver, D. Briscoe, M. Dangerfield, M. Gillings, and A. Beattie. 2004. Spatial scaling of microbial eukaryote diversity. Nature 432:747–750.

Hillebrand, H., F. Watermann, R. Karez, and U. G. Berninger. 2001. Differences in species richness patterns between unicellular and multicellular organisms. Oecologia 126(1): 114–124.

Horner-Devine, M. C., and B. J. M. Bohannan. 2006. Phylogenetic clustering and overdispersion in bacterial communities. Ecology 87:S100–S108.

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 B 271:113–122.

Horner-Devine, M. C., J. Green, and B. J. M. Bohannan. 2006. Patterns in prokaryotic biodiversity in prokaryote diversity: mechanisms and significance. Pages 19–38 *in* N. A. Logan, H. M. Lappin-Scott, and P. C. F. Oyston, editors. Prokaryote diversity: mechanisms and significance. Society for General Microbiology, Reading, UK.

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.

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

Martiny, J. B. H., et al. 2006. Microbial biogeography: putting microorganisms on the map. Nature Reviews Microbiology 4:102–112.

Orphan, V. J., K.-U. Hinrichs, W. Ussler III, C. K. Paul, L. T. Taylor, S. P. Sylva, J. M. Hayes, and E. F. Delong. 2001. Comparative analysis of methane-oxidizing archaea and sulfate-reducing bacteria in anoxic marine sediments. Applied and Environmental Microbiology 67:1922–1934.

Smith, V. H., B. L. Foster, J. P. Grover, R. D. Holt, M. A. Leibold, and F. deNoyelles, Jr. 2005. Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. Proceedings of the National Academy of Sciences (USA) 102(12):4393–4396.

Stone, L., and A. Roberts. 1990. The checkerboard score and species distributions. Oecologia 85(1):74–79.

Tringe, S. G., et al. 2005. Comparative metagenomics of microbial communities. Science 308:554–557.

van der Gast, C. J., A. K. Lilley, D. Ager, and I. P. Thompson. 2005. Island size and bacterial diversity in an archipelago of engineering machines. Environmental Microbiology 7(8): 1220–1226.

Webb, C. O. 2000. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. American Naturalist 156(2):145–155.

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.

APPENDIX

Breakdown of microbial data sets by taxon, habitat, method, and gene (*Ecological Archives* E088-083-A1).