

# Spatial scaling of functional gene diversity across various microbial taxa

Jizhong Zhou\*<sup>†‡</sup>, Sanghoon Kang\*<sup>†</sup>, Christopher W. Schadt<sup>†§</sup>, and Charles T. Garten, Jr.<sup>†</sup>

\*Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019; and Divisions of <sup>†</sup>Environmental Sciences and <sup>§</sup>Biosciences, Oak Ridge National Laboratory, Oak Ridge, TN 37923

Edited by James M. Tiedje, Michigan State University, East Lansing, MI, and approved March 25, 2008 (received for review September 25, 2007)

**Understanding the spatial patterns of organisms and the underlying mechanisms shaping biotic communities is a central goal in community ecology. One of the most well documented spatial patterns in plant and animal communities is the positive-power law relationship between species (or taxa) richness and area. Such taxa–area relationships (TARs) are one of the principal generalizations in ecology, and are fundamental to our understanding of the distribution of global biodiversity. However, TARs remain elusive in microbial communities, especially in soil habitats, because of inadequate sampling methodologies. Here, we describe TARs as gene–area relationships (GARs), at a whole-community level, across various microbial functional and phylogenetic groups in a forest soil, using a comprehensive functional gene array with >24,000 probes. Our analysis indicated that the forest soil microbial community exhibited a relatively flat gene–area relationship (slope  $z = 0.0624$ ), but the  $z$  values varied considerably across different functional and phylogenetic groups ( $z = 0.0475$ – $0.0959$ ). However, the  $z$  values are several times lower than those commonly observed in plants and animals. These results suggest that the turnover in space of microorganisms may be, in general, lower than that of plants and animals.**

canonical correspondence analysis | GeoChip | taxa–area relationship

The complex diversity of life is the most stunning feature of Earth and understanding such diversity and its distribution patterns at multiple spatial scales is a central issue in ecology (1, 2). Knowledge about the spatial distribution patterns of biodiversity are critical to deciphering the forces shaping and maintaining biodiversity (1) and are of practical importance for predicting species extinction risk because of loss of habitat (1) and for designing reserves to protect biodiversity (1). Therefore, the spatial distribution patterns of biodiversity have received a great deal of attention in macrobial ecology (1), but only more recently in microbial ecology (3–7).

The tendency that species richness increases with area [the species–area relationships (SARs), taxa–area relationships (TARs), or gene–area relationships (GARs)] is one of the few laws in ecology (8). Several mathematical equations have been used to describe TARs, but the power-law,  $S = cA^z$ , is one of the most commonly observed and used to describe taxa–area relationships (9). Here,  $S$  is the number of species,  $A$  is the area, and  $c$  is the intercept in log-log space. According to this formula, the species–area exponent,  $z$ , is a measure of the rate of species turnover across space. Such positive power-law relationships between species number and area are commonly observed in plant and animal communities. In macroorganisms, the  $z$  values are generally consistent across taxa and are often relatively close to a theoretically derived value of 0.25 (10) based on the canonical lognormal distributions of species.

Microorganisms are the most abundant and diverse group of life on our planet and play integral and often unique roles in the biogeochemical cycling of various materials and elements that are crucial to ecosystem functioning. Despite their importance, the spatial scaling of microbial diversity is poorly understood. Because of their unique biology, it has been long assumed that microorganisms have cosmopolitan distributions (11), which could lead to

fundamentally different biodiversity scaling relationships from those observed in plants and animals (11). Recently, a few studies have shown that microorganisms exhibit the power-law relationships between taxa richness and area based on measurements of microbial taxa defined by either morphological or a few molecular markers (e.g., 16S rRNA gene) (3–5); however, only very limited numbers of phylogenetic groups (e.g., *β-proteobacteria*) or communities at very coarse levels of resolution have been examined (5). Thus, the spatial scaling of microbial diversity at the whole community level across different functional and phylogenetic groups remains generally elusive and greatly understudied.

The slopes of the TARs for microorganisms vary substantially among different studies ( $z = 0.019$  to  $0.470$ ) (3–6, 11–14), especially in soils, which could be the most challenging habitats among all natural environments for microbiologists because of the microbial community size and the extent of species diversity. One study showed that the slope of the TARs in forest soils was barely noticeable ( $z = 0.03$ ) (3) and much lower than those in plants and animals, which is consistent with those from the majority of microbial studies (11). However, another study on forest soils recently showed the opposite trend ( $z = 0.42$  to  $0.47$ ), despite similar techniques used (6). Such variation in results could be due to the true differences in microbial spatial distributions but could also be compounded by the differences of the experimental design and analytical approaches used and/or by various sampling artifacts, such as under-sampling, unequal sampling, and taxonomic lumping (11, 15, 16) [see [supporting information \(SI\) Results and Discussion](#)]. Because microbial diversity is immense, it is likely that all studies have greatly under-sampled microbial diversity (4), especially when the dominant taxa are most widely distributed. Under-sampling could result in flat TARs (11). To minimize the artifact of under-sampling, more sampling effort is required to describe diversity in larger areas than smaller areas (11), which often leads to other artifacts related to unequal sampling. Such unequal sampling could lead to increases in diversity with area, even if the real community diversity remains unchanged (1, 11, 15). In addition, another potential sampling artifact could also occur with random sampling process. Generally, only small portions (< 1% species/populations) from a microbial community are randomly sampled for determining TARs, using conventional PCR-based molecular methods (5, 6, 17); thus, the probability of sampling the same portion of a community in various sampling events for determining TARs could be small given the size and diversity of soil communities. It is expected that overestimation of the slope of TARs would occur if different portions of a community are

Author contributions: J.Z. and S.K. contributed equally to this work; J.Z., C.W.S., and C.T.G. designed research; S.K., C.W.S., and C.T.G. performed research; S.K. contributed new reagents/analytic tools; S.K. analyzed data; and J.Z. and S.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>†</sup>To whom correspondence should be addressed. E-mail: jzhou@ou.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0709016105/DCSupplemental](http://www.pnas.org/cgi/content/full/0709016105/DCSupplemental).

© 2008 by The National Academy of Sciences of the USA

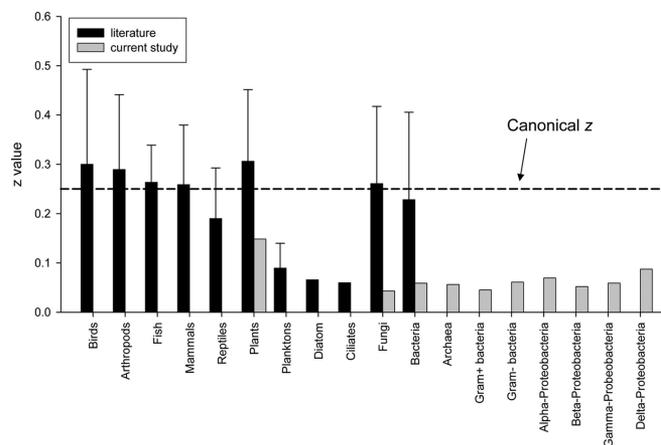


greater than that for the genes involved in nitrate reduction (0.0485) (Table 1). Also, the  $z$  value for fungi was 0.0475, which was significantly lower than those for Bacteria (0.0626) and Archaea (0.0748) (Tables S1 and S2). In addition, among *Proteobacteria*, the  $z$  value for  $\beta$ -*proteobacteria* (0.0519) was significantly lower than those for  $\alpha$ - (0.0662) and  $\gamma$ -*proteobacteria* (0.0644), indicating that, as a group, the  $\beta$ -*proteobacteria* may have a slower turnover in space than other *Proteobacteria* in the forest soil. This was consistent with the observation that the turnover in space of  $\beta$ -*proteobacteria* was lower than that of other bacteria sampled in salt marsh sediments (5). Finally, the  $z$  values for both fungi and Gram-positive bacteria were more comparable to each other and generally less than the  $z$  values estimated for the other organisms sampled. This latter difference could be related to the unique biology of spore formation in fungi and common soil bacterial genera, such as *Clostridia* and *Bacillus*, which may also rely on spores for dispersal.

Unlike traditional biogeography studies in plants and animals, individual genes rather than the whole organisms were detected for analysis in microbial biogeographical studies. Any genes could be subjected to various DNA recombination processes during evolution. Although using functional genes as biomarkers may provide greater taxonomic resolution than rRNA genes, it should be noted that functional gene biomarkers are more vulnerable to the effects of horizontal gene transfer (HGT), especially for those genes that are frequently plasmid-borne, such as those involved in metal resistance and organic contaminant degradation (19). The possible existence of HGT among these genes could result in underestimating true  $z$  values (5). To assess the potential impacts of HGT on GARs in this study, we calculated  $z$  values separately for different phylogenetic groups by simply excluding the genes that are plasmid-borne or potentially plasmid-borne, and/or excluding these genes involved in metal resistance and organic contaminant degradation because some of them are shown to be more vulnerable to HGT. Almost identical  $z$  values for different phylogenetic groups were obtained. This suggests that the impact of HGT on GARs was not a serious problem in our study.

Environmental heterogeneity and the differences in demographic processes (e.g., dispersal, colonization, speciation, and extinction) are important in determining TARs across different organisms (16, 20). To determine whether soil property variables affected microbial community composition, partial Mantel tests were performed. When the effects of geographic distance were held constant, partial Mantel tests indicated no significant correlations between the two datasets. Because using many unrelated individual variables can mask the signature of any significant variables in Mantel tests BIO-ENV and canonical correspondence analysis (CCA) were further performed to identify the abiotic factors important to community composition. Ammonium concentration, C/N ratio and fraction of C in particulate organic matter (fPOM-C) were identified as important factors, which were then used for constructing a Euclidean distance-based similarity matrix. Partial Mantel test revealed significant correlations among these three variables identified and community composition based on FGs ( $r_M = 0.293$ ,  $P < 0.05$ ) (Table S3). Similarly, partial Mantel tests revealed no significant correlations between geographic distance and all of the functional and phylogenetic groups except for the functional group involved in nitrate reduction, which exhibited a significant correlation with geographic distance ( $r_M = 0.169$ ,  $P < 0.05$ ) (Table S3).

Because partial CCA has been shown to be more appropriate to correctly partition the beta diversity among sites than partial Mantel test (21), a CCA-based variation partitioning analysis (VPA) (22) was further performed. Similar to the Mantel test, when the effects of geographic distance were held constant, no significant correlations were observed between the overall microbial community composition and the measured soil environmental variables. The reverse was true for geographic distance when soil environmental variables were held constant. To separate the effects of environ-



**Fig. 2.** A comparison of  $z$  values of macrobial and microbial taxonomic groups. Data were obtained from supplemental data of Drakare *et al.* (20) except those taxonomically less defined groups. Some recent data for microbial communities were also added. A total of 701 datasets were analyzed. The bars represented the average  $z$  values, and error bars represented 1 standard deviation. The data from other studies are presented as black bars; the data generated from this study are presented as grey bars.

mental soil variables and geographic distance, additional VPA was carried out based on the richness of all individual FGs. Although they were not statistically significant, about one fifth of the total microbial community composition variation could be explained independently by environmental heterogeneity (part a, 20.7%,  $P = 0.920$ ) and geographic distance (part c, 18.3%,  $P = 0.294$ ) (Fig. S3). A very small portion of the variation (part b, 5.8%) was contributed by both environmental heterogeneity and geographic distance, suggesting that the microbial community and environmental data have quite different spatial structuring. This is also supported by spatial correlogram analysis, which indicated that the overall FGS richness had a single patch at  $\approx 300$  m, whereas environmental variables had multiple patchinesses at  $\approx 50$  m and  $\approx 800$  m (Fig. S4). Similar to other studies (21), substantial amount (part d, 55.2%) of the microbial community variation could not be explained by the measured environmental variables and the geographical distance examined.

To obtain general insights on the spatial scaling of biodiversity across different organisms, the  $z$  values obtained in this study were compared with all available published data (701 datasets) (Fig. 2). Because the  $z$  values in these different studies were obtained with a variety of approaches at various spatial scales (e.g., contiguous vs. island habitats), it would be impractical to make exact comparisons at the detailed fine levels. Such comparisons would be especially more difficult between microorganisms and macroorganisms because the biogeography studies with microorganisms are often based on genes rather than individual organisms. If we assume that the difference of gene richness will reflect the difference of organism richness in microbial communities, the following coarse-level generalizations can be drawn. First, no overall significant differences of spatial scaling between plants and animals were observed. The mean  $z$  value for plants was  $0.306 (\pm 0.018)$ , whereas it was  $0.274 (\pm 0.017)$  for animals (Fig. 2), which are both relatively close to a theoretically derived value of 0.25 (10). Second, the turnover rate of microorganisms in space appears to be much lower than those of other organisms. The  $z$  values for microorganisms were generally  $< 0.1$  in contiguous habitats, which were three to four times less than those observed in higher plants and animals (Fig. 2). Additionally, overall, no clear-cut differences of spatial scaling between eukaryotic and prokaryotic microorganisms were observed (Fig. 2).

## Discussion

Microorganisms are very small, invisible to the naked eye, and their diversity is extremely high, with most (> 99%) microorganisms as-yet uncultured. Characterizing such vast diversity and understanding the mechanisms shaping it presents numerous obstacles. Because of the difficulty in obtaining quantitative experimental data, community-wide studies on the spatial and temporal dynamics of microbial community structure and activities in microbial ecology are most often conducted at a descriptive level rather than at a desired model-based predictive level (17). Because of this, it is often not clear whether existing ecological theories and models developed in macrobial ecology are applicable to microbial ecology (23). The development of microarray-based high throughput genomic technologies has the potential to enable microbial ecologists to test model-based ecological theories and hypotheses at the community-wide scales by allowing consistent and quantitative datasets to be generated on the large numbers of genes and samples that are required for addressing such questions. Thus, in this study, we have used a comprehensive functional gene array to examine the spatial patterns of microbial diversity in forest soil microbial communities.

Microarray hybridization-based approach has several advantages for examining TARs by eliminating or minimizing various sampling artifacts (see *SI Results and Discussion* for details). One of the main advantages is that the arrays contain tens of thousands of functional gene markers so that many microbial populations and functional groups can be simultaneously detected at the whole community-wide scale. In contrast to other molecular methods, the spatial distribution patterns of many individual functional genes in the same sample sets can be simultaneously obtained with the microarray-based hybridization approach, as illustrated in Table 1. Also, the information on microbial diversity based on many functional genes/groups could be more representative of the overall picture of microbial diversity at the whole community level than those based on a single gene (e.g., 16S rRNA gene). In addition, the microbial population diversity based on functional genes will be much lower than that based on 16S rRNA genes (*SI Results and Discussion*), and hence less sampling efforts will be required to survey the functional gene diversity of the microbial community of interest. As a consequence, the undersampling problem (11), which can be assessed by accumulation curves, could be ameliorated with the array hybridization-based detection approach outlined here. Finally, distance-decay approach (4, 5) uses relative comparisons of microbial community composition rather than richness to examine TARs. It also has advantages in ameliorating the undersampling problem because, it is unnecessary to fully characterize or sample a community for making relative comparisons, which could still give robust measures of spatial scaling parameters (5). By combining the array-based detection methods with such data analysis approaches, we believe that undersampling problem could be satisfactorily solved, if not eliminated.

Because of unique biology of microorganisms, their detection is often dependent on individual genes or other DNA fragments. Both phylogenetic (e.g., 16S rRNA, *gyrB*, *recA*, and ribosomal intergenic regions) and functional marker genes (e.g., *amoA*, *nirS*, *nirK*, *nifH*, *dsrAB*, and other biogeochemically important genes) are very useful for studying phylogenetic relationships among different organisms, for analyzing microbial community structure, and/or for monitoring the physiological status and functional activities of microbial populations and communities in natural environments (24). They can be used as molecular markers in examining spatial distribution patterns of microbial communities as demonstrated by the studies in refs. 4 and 5 and this study. Both phylogenetic and functional marker genes also have some distinct advantages in examining spatial distribution patterns of microbial

communities. Beyond the advantage of less diversity needed to be surveyed as mentioned above, another main advantage of the functional genes (e.g., *amoA*, *nirS*, *nirK*, *nifH*, and *dsrAB*) over 16S rRNA genes is that they can generally provide higher resolution in differentiating various microorganisms, and hence the taxonomic lumping problem (11) can be ameliorated. For instance, under the hybridization conditions of 50°C plus 50% formamide, the GeoChips can provide species/strain level resolution for analyzing microorganisms involved in various biogeochemical processes (25, 26), and hence the  $z$  values, closely approaching those in plants and animals, theoretically could be obtained. However, compared with 16S rRNA genes, most functional genes could provide less robust information on phylogenetic relationships among various organisms, especially because many functional genes are more vulnerable to horizontal gene transfer.

When molecular approaches are used for determining spatial distribution patterns of microbial communities, GARs, instead of traditional SARs, are obtained. Because prokaryotic taxa and species concepts are often poorly defined, using GARs could be more appropriate than using SARs or TARs. In this study, we used functional gene richness instead of “species” richness, to describe the spatial distribution patterns of the microbial community in relevance to area sizes. We hypothesize that the difference of functional gene richness as a whole will reflect the difference of “species” richness among different samples, and that functional genes are useful in evaluating spatial scaling of microbial communities across different taxa along with other phylogenetic marker genes (e.g., 16S rRNA genes). Thus, the  $z$  values determined based on many functional genes as a whole should, in theory, reflect the  $z$  values comparable to plants and animals. This is supported by the recent studies based on the average nucleotide identity (ANI) of the shared genes between two strains (27), which demonstrated that the ANI values of 94% corresponded to the traditional 70% DNA–DNA reassociation standard of the current species definition. This is also supported by our results showing that the 50-mer-based GeoChips can, given that stringent probe design and hybridization conditions are applied, provide species/strain level resolution for analyzing microorganisms involved in various biogeochemical processes (25, 26). Although great caution is also needed to make meaningful comparisons because the spatial distribution patterns of individual genes could be different from those of individual organisms because of the differences in evolutionary rates, gene copy numbers, and vulnerability to horizontal gene transfer of individual genes. However, we believe that more reliable comparisons can be made with the  $z$  values determined based on many functional genes as a whole as presented in this study. Further work is needed to validate this hypothesis.

The relationship between biodiversity and area is a central issue in ecology. Despite their ecological and biogeochemical importance, the TARs in microbial communities are poorly understood. Because of the unique biology of microorganisms (e.g., high dispersal rates, high functional redundancy, massive population sizes, rapid asexual reproduction, resistance to extinction, and horizontal gene transfer) (5, 11, 17), one might expect that the relationships of taxa richness to area could be quite different from those observed in macrobial ecology (5). Based on many different functional gene markers important to various biogeochemical, ecological, and environmental processes, our study demonstrates that microbial functional and phylogenetic groups exhibit TARs, although they vary considerably among functional/phylogenetic groups. Although the general values that affect the shape of these relationships may differ, our results strongly support the general claim that the TAR appears to be a universal law in biology (4, 5, 8)

The observed  $z$  values varied greatly across different studies that included different habitats for microorganisms (4–6, 12). Our results indicated the mean calculated  $z$  value were 0.0655 for different functional groups and 0.0640 for different phylogenetic

groups. These values might be expected to more accurately represent the  $z$  values in soil microorganisms. Using 16S rRNA gene-based PCR cloning and sequencing approaches, Horner-Devine *et al.* (5) estimated that the  $z$  values were 0.04 for bacteria and 0.019 for  $\beta$ -proteobacteria at the taxonomic resolution of 99% sequence identity. The  $z$  values estimated in our study are  $\approx 2$ –3 times greater for bacteria (0.0626) and  $\beta$ -proteobacteria (0.0519) than those based on 16S rRNA genes. The main reason for this could be that the taxonomic resolution based on the functional genes on the arrays is higher than that based on 16S rRNA gene (28). However, our results and those of others suggest that  $z$  values are typically  $< 0.1$  for microorganisms, which are still much less than those typically observed in plants and animals (5, 20).

The TARs for fungi were recently studied in four different arid land systems, using ARISA (automated ribosomal intergenic spacer analysis). The estimated  $z$  values were 0.0175, 0.0235, 0.0315, and 0.038 for four different systems individually, whereas the  $z$  value across all of the data from the four systems combined was 0.074 (4). Similar to the results observed for bacterial groups, the calculated  $z$  value (0.0475) for fungi in our study was slightly greater, but it still falls within the general range of  $z$  values observed in these systems. Because different methods were used with different ecosystems, direct comparison of  $z$  values is difficult. However, as the authors indicated, the ARISA-defined OTU could also underestimate the number of fungal species, which could lead to lower  $z$  values. In addition, the estimated  $z$  value is substantially lower than the  $z$  values (0.20–0.23) for ectomycorrhizal fungi (7).

Very few studies have examined the taxa–area relationships in forest soils. The calculated  $z$  values in our study were more or less consistent with that ( $z = 0.03$ ) obtained from 98 forest soil samples from many different sites (3), but they were much lower than those in tropical deciduous forest soils ( $z = 0.42$  for the hilltop soil and 0.47 for the slope soil), using 16S rRNA genes-based T-RFLP analysis (6). The latter values are higher than the  $z$  values reported for vertebrates and sea invertebrates (1). We are not sure what causes such differences, but the sampling artifacts and differences in experimental design could be one of the main reasons. Because our experimental design is more consistent with typical biogeographical studies (5) and many functional genes were simultaneously examined, we believe that the  $z$  values presented in this study could be more representative of the spatial distribution patterns of microbial communities in forest soils.

Both environmental heterogeneity and the differences in demographic processes play important roles in determining TARs across different organisms (16, 20). The factors controlling TARs in the forest soil could be very complicated. It appears that environmental heterogeneity and geographic distance have very little influences on the diversity patterns of the microbial communities in these soils because no significant relationships were detected for most of the functional gene groups examined by using both partial Mantel test and partial CCA. There are several possible explanations for these observations. First, similar to several other studies (21), substantial amount of the variations of microbial community compositions as measured based on functional genes could not be explained by the measured environmental variables and spatial structure. A fair amount of the unexplained variation could be due to the local effects of unmeasured biotic (e.g., competitions and trophic interactions) or abiotic (e.g.,  $O_2$  level within soil aggregates and labile C pool) controlling variables or the missed description of spatial structures (see below) (22). Second, the spatial scale used for sampling in this study may not be suitable for discerning the influences of environmental heterogeneity and geographic distance on microbial community diversity patterns. A larger spatial scale (on the order of tens of thousands of kilometers) may be more appropriate for revealing the impacts of geographic distance on microbial communities (16), while smaller spatial scales ( $< 1$  m<sup>2</sup> level or even the soil aggregate particle level) may be needed to discern the influences of environmental heterogeneity. The data on

soil properties in our study showed few significant differences in variation between small scales ( $\approx 1$  m) and large scales ( $\approx 1$  km), suggesting that variation at smaller scales in these soils may be just as great as that at larger scales (29). Natural selection on microbial populations by soil chemical and physical factors most likely occurs at even smaller scales, such as soil particles and aggregates. In addition, it is not impossible that a part of the microbial community variation observed could be a result of ecologically neutral processes of diversification through random ecological drift (30). The genetic diversity of functional genes generated by neutral processes would provide little correlation to the environmental and spatial variables examined. Further studies are needed to examine both biotic and abiotic mechanisms in shaping microbial community composition and structure.

It should be noted that microarray-based hybridization could also underestimate the  $z$  values if the probes on the arrays do not well represent the diversity of the genes of interest in a given microbial community. If many genes of interest are not represented on the arrays and if these genes exhibit high spatial variability, the estimated  $z$  values based on array hybridization could be underestimated. One solution to this problem is to use metagenomic approaches to characterize the molecular diversity of the genes of interest in a community. Then the probes can be designed from any new gene sequences of interest for microarray fabrication. This will ensure that the designed arrays well represent the molecular diversity of the communities of interest if sufficient sequence information is available. Because it has many probes from different functional genes, the GeoChips used in this study could detect reasonably well some intensively studied functional genes although continuous improvement with higher gene coverage is always needed. In addition, the knowledge gained from this study on spatial scaling of microbial diversity could be only applicable at the meter scale used in this study, but may be not hold at smaller (e.g., micrometer or soil aggregate scales) or larger (e.g., tens of thousand of kilometers) scales of study.

In conclusion, understanding spatial scaling of biodiversity is a central question in ecology. Using the comprehensive GeoChip, our results revealed that the forest soil microbial communities exhibit gene–area relationships at the whole community level across various functional and phylogenetic groups. Because several artifacts in determining GARs for microorganisms can be minimized or eliminated by using the microarray hybridization-based detection approach, the estimated slopes of GARs are likely to more realistically reflect the spatial scaling of biodiversity in microbial communities. The determined slopes of GARs across different functional and phylogenetic groups in the forest soil were  $< 0.1$ , which is much lower than those in plants and animals. However, although the scaling of these relationships appear to be much different from that macroorganismal communities, our results strongly support the hypothesis that the taxa–area relationship is a universal law in biology (4, 5, 8) applies to soil microorganisms. It is also expected that this array-based approach will provide a powerful tool for studying microbial biogeography under various conditions in various types of habitats, including soils, sediments, groundwaters, wastewater treatments, and animal guts, and provide insights on microbial spatial patterns and the underlying controlling mechanisms.

## Materials and Methods

**Sampling and Soil Analysis.** Two perpendicular 1-km transects crossing at the center of each other were located in a deciduous forest at the Oak Ridge Reservation (Oak Ridge, TN). Sampling locations were placed at the center point and at 1-, 5-, 10-, 50-, 250-, and 500-m distances from the center point along transects in all four directions (*SI Materials and Methods* for details). Four soil cores (15 cm deep) around each sampling point were randomly collected for chemical and molecular analysis. A total of 23 geochemical variables were measured and derived. Additionally, at each sampling point, the predominant overstory tree species within 10 m of the center point were recorded.

**DNA Extraction and Microarray Hybridization.** Community DNA extraction, community DNA amplification (100 ng of purified community DNA in triplicate), probe labeling, microarray hybridization, scanning, and image processing were performed as described in refs. 31 and 32 (see *SI Materials and Methods* for details). The GeoChip used in this study contained 24,243 oligonucleotide probes targeting >150 functional groups of >10,000 genes essential to the biogeochemical cycles of carbon, nitrogen, phosphorus, and sulfur along with metal resistance, metal reduction and organic contaminant degradation (18). Microarray hybridization was carried out overnight at 50°C in 50% formamide, in which specific hybridization can be achieved for probes with sequence identity of <90% to the target sequences (18, 25), even up to 98% similar to the target sequences for some probes (25). Thus, the GeoChip hybridization could provide species-strain level of resolution (26).

**Data Processing and Analysis.** Microarray data preprocessing (e.g., removing poor and outlying spots, normalization) was carried out as described in refs. 18 and 31 (see *SI Materials and Methods* for details). The normalized hybridization data for individual functional gene sequences was then reorganized based on functional genes, such as *nifH* and *nirS*, and functional groups, such as nitrification, denitrification for estimating  $z$  values.

Richness–area modeling was fitted by linear regression in log space for the observed gene richness ( $S_{\text{obs}}$ ) of seven nested areas (0.02, 2, 50, 200, 2,500, 125,000, and 500,000 m<sup>2</sup>). The slope ( $z$  value) was tested by an one-sample  $t$  test between original slope and a mean of bootstrapped slopes by random pairing of the original set (10,000 times with replacement) because of lack of independence

from the nested sampling design (5) (*SI Materials and Methods*). The significance comparison of  $z$  values among different estimations was also achieved by bootstrapping (1,000 times), followed by pairwise  $t$  test with Bonferroni correction.

Sorensen index was used to construct community similarity matrices of both microorganisms and plants for analysis with BIO-ENV and Mantel test, whereas Euclidean distances were used to construct similarity matrices for geographic distance and soil properties. Partial Mantel tests and partial CCA were used to determine the effects of environmental heterogeneity and plant tree composition on microbial community composition by controlling geographic distance as constant and vice versa (5). BIO-ENV and CCA were also used to identify the abiotic factors most important to community composition. The partial CCA used for variation partition was carried out as described in ref. 22.

**ACKNOWLEDGMENTS.** We thank Jessica L. Green, Yiqi Luo, and Brendan Bohannan for comments on the manuscript and two anonymous reviewers for their invaluable comments and suggestions. This work was supported by The United States Department of Energy under the Carbon Sequestration program (as part of the consortium on research to enhance Carbon Sequestration in Terrestrial Ecosystems) and by the University of Oklahoma Research Foundation. The GeoChip used in this study was supported by the Genomics:GTL program through the Virtual Institute of Microbial Stress and Survival (<http://vimss.lbl.gov>), the Environmental Remediation Science Program, and Biotechnology Investigation—Ocean Margins Program of the Office of Biological and Environmental Research, Office of Science. Oak Ridge National Laboratory is managed by University of Tennessee UT-Battelle LLC for the Department of Energy under Contract DE-AC05-00OR22725.

- Rosenzweig ML (1995) *Species diversity in space and time* (Cambridge Univ Press, Cambridge, UK).
- Harte J, Conlisk E, Ostling A, Green JL, Smith AB (2005) A theory of spatial structure in ecological communities at multiple spatial scales. *Ecol Monogr* 75:179–197.
- Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci* 103:626–631.
- Green JL, et al. (2004) Spatial scaling of microbial eukaryote diversity. *Nature* 432:747–750.
- Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM (2004) A taxa–area relationship for bacteria. *Nature* 432:750–753.
- Noguez AM, et al. (2005) Microbial macroecology: Highly structured prokaryotic soil assemblages in a tropical deciduous forest. *Glob Ecol Biogeogr* 14:241–248.
- Peay KG, Bruns TD, Kennedy PG, Bergemann SE, Garbelotto M (2007) A strong species–area relationship for eukaryotic soil microbes: Island size matters for ectomycorrhizal fungi. *Ecol Lett* 10:470–480.
- Lawton JH (1999) Are there general laws in ecology? *Oikos* 84:177–192.
- Martin HG, Goldenfeld N (2006) On the origin and robustness of power-law species–area relationships in ecology. *Proc Natl Acad Sci* 103:10310–10315.
- Southwood TRE, May RM, Sugihara G (2006) Observations on related ecological exponents. *Proc Natl Acad Sci* 103:6031–6033.
- Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. *Trends Ecol Evol* 21:501–507.
- Bell T, et al. (2005) Larger islands house more bacterial taxa. *Science* 308:1884.
- Smith VH, et al. (2005) Phytoplankton species richness scales consistently from laboratory microcosms to the world's oceans. *Proc Natl Acad Sci* 102:4393–4396.
- Woodcock S, Curtis TP, Head IM, Lunn M, Sloan WT (2006) Taxa–area relationships for microbes: The unsampled and the unseen. *Ecol Lett* 9:805–812.
- Cam E, et al. (2002) Disentangling sampling and ecological explanations underlying species–area relationships. *Ecology* 83:1118–1130.
- Martiny JBH, et al. (2006) Microbial biogeography: Putting microorganisms on the map. *Nat Rev Microbiol* 4:102–112.
- van der Gast CJ, Lilley AK, Ager D, Thompson IP (2005) Island size and bacterial diversity in an archipelago of engineering machines. *Environ Microbiol* 7:1220–1226.
- He Z, et al. (2007) GeoChip: A comprehensive microarray for investigating biogeochemical, ecological, and environmental processes. *ISME J* 1:67–77.
- Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 3:711–721.
- Drakare S, Lennon JJ, Hillebrand H (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecol Lett* 9:215–227.
- Ramette A, Tiedje JM (2007) Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proc Natl Acad Sci* 104:2761–2766.
- Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology* 73:1045–1055.
- Prosser JI, et al. (2007) The role of ecological theory in microbial ecology. *Nat Rev Microbiol* 5:384–392.
- Harayama S, Kasai, H (2006) in *Molecular Identification, Systematics, and Population Structure of Prokaryotes*, ed. Stackebrandt, E. (Springer, New York, NY), pp 105–140.
- Liebich J, et al. (2006) Improvement of oligonucleotide probe design criteria for functional gene microarrays in environmental applications. *Appl Environ Microbiol* 72:1688–1691.
- Tiquia SM, et al. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* 36:664–675.
- Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci* 102:2567–2572.
- Stackebrandt E, Goebel BM (1994) Taxonomic note: A place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* 44:846–849.
- Garten CT, Jr, Kang S, Brice DJ, Schadt CW, Zhou J (2007) Variability in soil properties at different spatial scales (1 m to 1 km) in a deciduous forest ecosystem. *Soil Biol Biochem* 39:2621–2627.
- Ramette A, Tiedje JM (2007) Biogeography: An emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* 53:197–207.
- Wu L, Liu X, Schadt CW, Zhou J (2006) Microarray-based analysis of sub-nanogram quantities of microbial community DNAs using Whole Community Genome Amplification (WCGA). *Appl Environ Microbiol* 72:4931–4941.
- Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62:316–322.