Biogeographic patterns of soil diazotrophic communities across six forests in the North America

QICHAO TU,*† YE DENG,‡ QINGYUN YAN,† LINA SHEN,† LU LIN,* ZHILI HE,† LIYOU WU,† JOY D. VAN NOSTRAND,† VANESSA BUZZARD,§ SEAN T. MICHALETZ,§ BRIAN J. ENQUIST,§** MICHAEL D. WEISER,†† MICHAEL KASPARI,†††‡ ROBERT B. WAIDE,§§ JAMES H. BROWN,§§ and JIZHONG ZHOU†¶¶***

*Department of Marine Sciences, Ocean College, Zhejiang University, Zhejiang 310058, China, †Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK 73019, USA, ‡Research Center for Eco-Environmental Science, Chinese Academy of Sciences, Beijing 100085, China, §Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA, ¶Earth and Environmental Sciences Division, Los Alamos National Laboratory, MS J495, Los Alamos NM 87545, USA, **The Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501, USA, ††Department of Biology, EEB Graduate Program, University of Oklahoma, Norman, OK 73019, USA, ‡‡Smithsonian Tropical Research Institute, Balboa, 0843-03092, Republic of Panama, §§Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA, ¶¶State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China, ***Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94270, USA

Correspondence: Jizhong Zhou, Fax: +1 405-325-7552; E-mail: jzhou@ou.edu

Abstract

Soil diazotrophs play important roles in ecosystem functioning by converting atmospheric N_2 into biologically available ammonium. However, the diversity and distribution of soil diazotrophic communities in different forests and whether they follow biogeographic patterns similar to macroorganisms still remain unclear. By sequencing *nifH* gene amplicons, we surveyed the diversity, structure and biogeographic patterns of soil diazotrophic communities across six North American forests (126 nested samples). Our results showed that each forest harboured markedly different soil diazotrophic communities and that these communities followed traditional biogeographic patterns similar to plant and animal communities, including the taxa–area relationship (TAR) and latitudinal diversity gradient. Significantly higher community diversity and lower microbial spatial turnover rates (i.e. *z*-values) were found for rainforests (~0.06) than temperate forests (~0.1). The gradient pattern of TARs and community diversity was strongly correlated ($r^2 > 0.5$) with latitude, annual mean temperature, plant species richness and precipitation, and weakly correlated ($r^2 < 0.25$) with pH and soil moisture. This study suggests that even microbial subcommunities (e.g. soil diazotrophs) follow general biogeographic patterns (e.g. TAR, latitudinal diversity gradient), and indicates that the metabolic theory of

ecology and habitat heterogeneity may be the major underlying ecological mechanisms shaping the biogeographic patterns of soil diazotrophic communities.

Introduction

Biologically available nitrogen (nitrate, ammonia and organic nitrogen) is a limited nutrient of crucial importance for ecosystem function in both agricultural and natural ecosystems (Vitousek & Howarth 1991). Reduction in atmospheric N_2 to ammonium via biological nitrogen (N) fixation is the major pathway by which atmospheric N_2 enters the biosphere, contributing about 128 Tg N per year to terrestrial ecosystems (Galloway *et al*. 2004). The *nifH* gene family, which encodes the nitrogenase reductase subunit, is widely used as a molecular marker to analyse diazotrophic microbial communities (Izquierdo & Nüsslein 2006; Moisander *et al*. 2006; Mohamed *et al*. 2008; Hsu & Buckley 2009; Zehr 2011; Groszkopf *et al*. 2012; Wang *et al*. 2013; Berthrong *et al*. 2014; Collavino *et al*. 2014). Both symbiotic and free-living diazotrophs make significant contributions to the Earth's N budget (Cleveland *et al*. 1999; Zehr 2011; Groszkopf *et al*. 2012). Although N_2-fixing communities have been studied in many different ecosystems, the extensive diversity and complex community structure of these communities have not been fully explored in soil ecosystems. More importantly, although much work has been done to assess the distribution of symbiotic N_2-fixing *Frankia* spp. (Benson *et al*. 2004; Benson & Dawson 2007; Franche *et al*. 2009), biogeographic- and macroecological-scale patterns of soil diazotrophic communities have rarely been investigated, until recently (Shay *et al*. 2015). Unlike many other microbial communities, soil diazotrophic communities are both free-living and symbiotic, and so may differ greatly in their biogeographic patterns. Thus, understanding biogeographic patterns is of crucial importance to exploring the underlying mechanisms controlling the structure and diversity of soil diazotrophic communities.

Biogeographic patterns of macro- and microbial communities, including the species–area relationship and geographic diversity gradient, are important ecological characteristics in understanding community diversity, structure and distribution. Macroorganisms, including plants and animals, are known to follow clear species–area relationships and latitudinal diversity gradient patterns. The small size, high abundance and as yet uncultivated status of the majority of microorganisms may lead to biogeographic-scale patterns radically different from macroorganisms. For example, microorganisms may disperse further and faster, resulting in almost nonexistent dispersal limitation and cosmopolitan distributions (Finlay 2002; Green & Bohannan 2006). Recent work is equivocal about similarities between the biogeographic-scale patterns of microbial and macrocommunities

(Green *et al*. 2004; Horner-Devine *et al*. 2004; Fierer & Jackson 2006; Follows *et al*. 2007; Fuhrman *et al*. 2008; Zhou *et al*. 2008; Ghiglione *et al*. 2012; Tedersoo *et al*. 2014). Notably, such disagreement on microbial biogeography could be due to a question of scale as implicated in a recent study (Shay *et al*. 2015), as different biogeographic properties may play different roles at regional, continental and global scales. Taxa–area relationships (TARs) of microbial communities were observed in many studies and were shown to have slopes substantially different from plant and animal communities (Martiny *et al*. 2006). However, it is uncertain whether microbial communities show latitudinal diversity gradients. In soil ecosystems, some studies have suggested a pH diversity gradient for terrestrial soil microbial communities, whereas an apparent latitudinal diversity gradient was not observed (Fierer & Jackson 2006; Griffiths *et al*. 2011). In marine ecosystems, studies have shown a traditional latitudinal diversity gradient for planktonic bacteria (decreasing diversity with increasing latitude) (Pommier *et al*. 2007; Fuhrman *et al*. 2008), but another showed increasing bacterial diversity with increasing latitudes (Ladau *et al*. 2013).

In this study, we hypothesize that (i) different soil microbial communities would be present in different types of forests harbouring divergent plant species and environmental characteristics; and (ii) similar to macroorganisms and other microbial communities, soil diazotrophs will also follow ecological principles such as taxa–area relationships and biogeographic diversity gradients, but with different patterns (e.g. different TAR slopes) due to the unique habitat of soil diazotrophs. To test these hypotheses, we quantified the diversity of diazotrophic communities within and between six forests to describe the taxa–area relationships as well as biogeographic (e.g. latitude and pH) diversity gradients by sequencing *nifH* gene amplicons. This study improves our understanding of the biogeographic patterns of soil diazotrophs and provides evidence that soil diazotrophic communities in forest soils also follow biogeographic diversity gradient patterns.

Materials and methods

Site description, experimental design and sampling

Six forest sites were selected for this study, including H.J. Andrews (AND, coniferous forest, 44°12′44.2″N, 122°15′19″W), Coweeta (CWT, deciduous forest, 35°3′37.2″N, 83°25′49.02″W), Harvard (HFR, deciduous forest, 42°32′16.08″N, 72°11′24″W), Luquillo (LUQ, tropical rainforest, 18°18′N, 65°48′W), Niwot Ridge (NWT, alpine tundra, 39°59′24″N, 105°22′48″W) and Barro Colorado Island (BCI, tropical rainforest, 9°09′N, 79°51′W). The selected sites provide variation in ecosystem types from boreal to tropical forest, across a latitudinal gradient from 9 to 44°N. The sites are characterized by considerable variations in: average annual temperature (from 2.5 to 25.7 °C), plant species richness (from 5 to 263 tree species),

annual precipitation (from 550 to 3460 mm), pH (from 3.41 to 6.63), and soil moisture (from 5.6% to 64.63%; see Fig. S1, Supporting information).

At each of the six forest sites across the latitudinal gradient, we used a nested sampling design to survey the background pools of regional microbial diversity. At each site, we collected and homogenized nine surface soil cores (~10 cm depth, Oakfield Apparatus Company model HA) from each of 21 individual square metre plots. The 21 1-m^2 plots were laid out in a cross-pattern (Fig. S1, Supporting information) with plots located at 1, 10, 50, 100 and 200 m in each cardinal direction from the central 1-m^2 plot.

Bulk soil samples were taken in the summer of 2012 for microbial community analysis. Soils were kept on ice in the field, then at −20 °C (LUQ, CWT, AND, NWT) or −80 °C (BCI, HFR) until shipped overnight on dry ice to the Institute for Environmental Genomics at the University of Oklahoma for DNA extraction, PCR amplification and sequencing.

Metadata collection

Plant species were surveyed using a modified 'Gentry plot' methodology whereby five 0.1-ha Gentry plots were established by B.J. Enquist, V. Buzzard and S. Michaletz *et al*. within the 25-ha plot within each site (Gentry 1982). Mean annual temperature and average annual precipitation were calculated from hourly data collected from onsite weather stations. The soil moisture was measured by putting 1.5 g soil into 65 °C oven until a constant weight was reached. The percentage of the original weight loss after oven drying was calculated as the soil moisture content (%). Soil pH was measured in a soil suspension with a soil: water ratio of 1:2.5 (weight: volume) using a standard protocol described previously (Ministry of Agriculture 1986). The soil C and N contents were measured by a LECO TruSpec Carbon and Nitrogen Analyzer (LECO Corporation, St. Joseph, MI) in the Soil, Water and Forage Analytical Laboratory at the Oklahoma State University (Stillwater, OK). In the same analytical laboratory, the soil NH_4^+ and NO_3^- contents were extracted from the soils with 1 M KCl using a standard protocol described previously (Bremner 1965) and measured by a Lachat QuikChem 8500 series 2 instrument (Lachat, Loveland, CO). More detailed information about this project and metadata collection can be found at http://macroeco.lternet.edu/.

DNA extraction, PCR amplification and sequencing

Soil DNA was extracted by freeze-grinding mechanical lysis as described previously (Zhou *et al*. 1996) and was purified using a low-melting agarose gel followed by phenol extraction for all 18 soil samples collected. DNA quality was assessed by 260/280 and 260/230 nm ratios using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE), and final soil DNA concentrations were quantified by PicoGreen (Ahn *et al*. 1996) using a FLUOstar Optima (BMG Labtech, Jena, Germany).

The PolF/R primer set (Pol115F, TGCGAYCCSAARGCBGACTC, and Pol457R, ATSGCCATCATYTCRCCGGA), which targets a 302-bp sequence of *nifH* gene,

was used for PCR amplification. A phasing amplicon sequencing approach (Wu *et al*. 2015) was used. A two-step PCR was performed to avoid extra PCR bias that could be introduced by the added components in the long primers used for PCR library preparation. In the two-step PCR method, the first-round PCR was carried out in a 50 µl reaction containing 5 µl 10 × PCR buffer II (including dNTPs), 0.5 U high-fidelity AccuPrime™ Taq DNA polymerase (Life Technologies), 0.4 µM of both forward and reverse target-only primers, and 10 ng soil DNA. Thermal cycling conditions were as follows: initial denaturation at 94 °C for 1 min, followed by 10 cycles of 94 °C for 20 s, 62 °C for 25 s and 68 °C for 45 s, with a final extension at 68 °C for 10 min. The triplicate products from the first-round PCR were combined, purified with an Agencourt® AMPure XP kit beads (Beckman Coulter, Beverly, MA, USA), eluted in 50 µl water and aliquoted into three new PCR tubes (15 µl each). The second-round PCR used a 25 µl reaction containing 2.5 µl 10 × PCR buffer II (including dNTPs), 0.25 U high-fidelity AccuPrime™ Taq DNA polymerase (Life Technologies), 0.4 µM of both forward and reverse phasing primers, and 15 µl of the first-round purified PCR product. The amplifications were carried out using the above programme with 20 cycles. Positive PCR products were confirmed by agarose gel electrophoresis. PCR products from triplicate reactions were combined and quantified with PicoGreen.

PCR products from samples to be sequenced in the same MiSeq run were pooled at equal molality. The pooled mixture was purified with a QIAquick Gel Extraction Kit (QIAGEN Sciences, Germantown, MD, USA) and requantified with PicoGreen. Sample libraries for sequencing were prepared according to the MiSeq™ Reagent Kit Preparation Guide (Illumina, San Diego, CA, USA) as described previously (Caporaso *et al*. 2012). For sequencing, a 500-cycle v2 MiSeq reagent cartridge (Illumina) was thawed for 1 h in a water bath, inverted ten times to mix the thawed reagents and stored at 4 °C for a short time until use. Sequencing was performed for 251, 12 and 251 cycles for forward, index and reverse reads, respectively.

Sequence processing and data analysis

Quality filtering, chimera removal and OTU clustering were carried out using the UPARSE pipeline (Edgar 2013), which is a recently developed approach that identifies highly accurate OTUs from amplicon sequencing data. Only the reads with perfectly matched barcodes were kept for further analysis. Forward and reverse reads with at least 50-bp overlap and fewer than 5% mismatches were combined into a single sequence by the FLASH program (Magoč & Salzberg 2011). Combined sequences with expected errors >0.5 or <250 bp were discarded. The FRAMEBOT program (Wang *et al*. 2013) was used to correct potential frameshifts caused by sequencing errors and only reads whose translated proteins mapped to reference *nifH* protein sequences with >30% identity were retained. The reads were then dereplicated, sorted and clustered into candidate OTUs with an identity cut-off of 0.94; the average nucleotide identity was used for microbial species definition in the postgenomic era (Konstantinidis & Tiedje 2005). Chimeric OTUs were then

identified and removed using uchime by searching against the *nifH* reference sequences maintained and curated by Zehr *et al*. (http://pmc.ucsc.edu/~wwwzehr/research/database/, 4 April 2014) (Zehr *et al*. 2003). Finally, qualified reads were mapped to representative OTU sequences for relative abundance calculation.

Taxonomic assignment for *nifH* OTUs was carried out by searching representative sequences against reference *nifH* sequences with known taxonomic information. A minimum global identity cut-off of 40% was used to filter searching results. Taxonomic assignment for *nifH* OTUs was performed by an in-house PERL script based on the best hits with the highest global identity, after searching against a curated database containing only *nifH* reference sequences with taxonomic information. The lowest common ancestor algorithm as described in the MEGAN (Huson *et al*. 2007) program was used for taxonomic assignment. Taxonomic information at the genus level or higher was assigned. Principle coordinate analysis (PCoA), Bray–Curtis dissimilarity analysis, canonical correspondence analysis (CCA) and statistical analysis including ANOSIM, MRPP and ADONIS were carried out by the VEGAN package in R (Oksanen *et al*. 2007). OTU relative abundance profiles generated from a random resampling effort of 8000 sequences per sample were used for statistical analysis. Species richness and rarefaction analysis were carried out by the MOTHUR package (Schloss *et al*. 2009), with a rarefaction analysis of 1000 bootstrap random sampling iterations and 0.1% incremental resampling effort.

Results

Summary of nifH amplicon sequencing

A total of 3 754 414 sequences were obtained for the 126 samples. After quality trimming, frameshift correction and chimera removal, 3 051 192 reads were clustered into 12 265 *nifH* OTUs at 94% identity cut-off, of which only 425 OTUs were detected in all six forests. The number of sequences in each sample ranged from 8420 to 43 271 (24 216 on average), resulting in 217 to 1904 OTUs (796 on average) per sample. A random resampling effort of 8000 sequences per sample was performed for further statistical analysis.

Distinct diazotrophic community structure, diversity and composition across the six forests

The soil diazotrophic community structure was significantly ($P < 0.001$) different across the six forests (Fig. 1A, Table S1, Supporting information). The overall community structure of soil diazotrophs was analysed by PCoA based on *nifH* OTU relative abundance profiles. Soil samples from different ecosystems were clearly separated from each other as a result. Rainforest soil samples were separated from temperate ecosystem samples by the second principle coordinate. NWT and HFR soil diazotrophic communities were clustered more closely together, indicating they were more similar. Such distinct soil diazotrophic community structures across the six forests

were also confirmed by three different nonparametric multivariate dissimilarity analysis methods, including ADONIS (permutation multivariate analysis of variance), ANOSIM (analysis of similarities) and MRPP (multiresponse permutation procedure), with P-values ≤ 0.001 (Table S1, Supporting information).
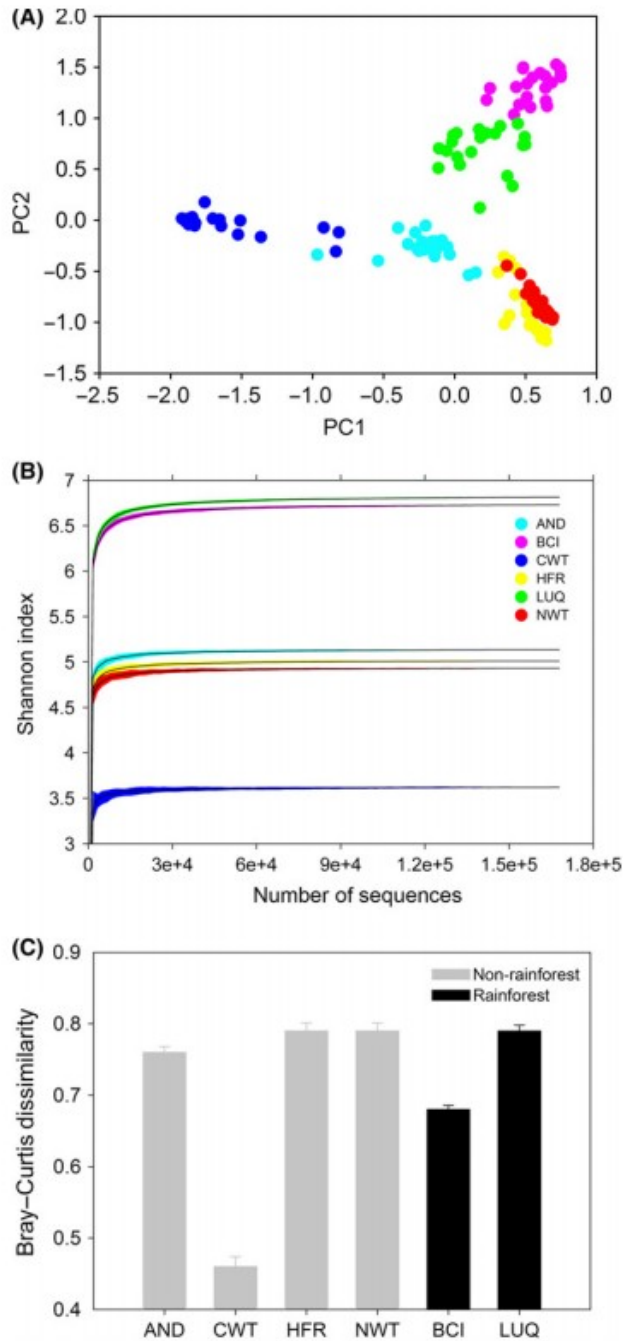
Fig. 1 The community structure and diversity of soil diazotrophs. (A) Principle coordinates of *nifH* communities. Distinct community structures could be observed between different forest ecosystems. (B) The Chao1 richness of *nifH* OTUs. Rainforest ecosystems harbour significantly higher OTU richness than temperate forest ecosystems. Shaded area indicates 95% confidence interval. (C) Within-site beta diversity of *nifH* communities. No significant differences were observed between rainforest and temperate forest ecosystems, except the Coweeta forest, which had significantly lower beta diversity than the other five ecosystems.

We used the Shannon index to estimate the diversity of soil diazotrophic communities across these six forests (Fig. 1B). Rainforest ecosystems such as LUQ and BCI harboured significantly higher *nifH* diversity than temperate forest ecosystems that no overlap of rarefaction curves was found at 95% confidence intervals at all sequencing depths we randomly resampled. With 21 samples and a sequencing depth of ~160 000 per site, the Shannon diversity for rainforests was ~6.8, whereas the diversity for temperate forests was between ~3.6 and 5.1. The CWT forest had the lowest diversity (i.e. ~3.6) among the six forests. At the sample level, similar trends were observed for both Chao1 richness and Shannon diversity index, with rainforests generally harbouring higher diazotrophic diversity than temperate forests, although some variation was observed for a few samples (Fig. S2, Supporting information). Notably, no significant differences were observed for the within-site beta diversity between the rainforests and temperate forests, except for CWT, in which a much lower beta diversity was found than for the other five ecosystems (Fig. 1C).

The phylogenetic and taxonomic composition of the soil diazotrophic communities across the six forests was also characterized (Fig. S3, Supporting information). Phylogenetic types of *nifH* genes were defined according to Young (2005). All of these ecosystems were dominated by type B ('bacterial' Mo-dependent nitrogenase) *nifH*-containing microorganisms, with ~91.5% to 99.5% total sequences per site. Only ~0.5% to 8.5% of type C ('clostridial' Mo-dependent nitrogenase) *nifH* genes were found in each forest soil (Fig. S3A, Supporting information). The taxonomic composition of the detected soil diazotrophs was also markedly different across different forests (Fig. S3B, Supporting information). Diazotrophic microorganisms belonging to Alphaproteobacteria were dominant in all forests, followed by Actinobacteria, Cyanobacteria, Beta-, Delta- and Gammaproteobacteria.

Dominant diazotrophs were distinctly different across the six forests

To assess if the samples from different ecosystems contained different soil diazotrophs at the OTU/species level, the dominant *nifH* OTUs were extracted and analysed. Here, *nifH* OTUs present in more than 50% of the samples and with a >0.5% relative abundance were considered dominant OTUs for each site. Completely different dominant *nifH* OTUs were found in the different forests (Fig. S4, Supporting information). For example, at HRF, *nifH* OTUs corresponding to *Rhodopseudomonas palustris*, *Methylocystis echinoides*, *Desulfovibrio sp*. U5L, *Rhodomicrobium vannielli* and *Azorhizobium caulinodans* species assigned as their closest taxonomic groups predominated. At AND and CWT, groups of *nifH* OTUs assigned to uncultured N$_2$-fixing bacterium C20 were dominant. At NWT, *nifH* OTUs closely related with *Bradyrhizobium sp*. IRBG 228, uncultured N$_2$-fixing bacterium B13 and *Methylocystis* species were dominant. In contrast, rainforests LUQ and BCI did not have an obviously dominant group of *nifH* OTUs, possibly due to the high diversity of soil diazotrophic communities at these sites.

Importance of diazotrophic diversity to soil biologically available nitrogen

To see whether soil diazotrophic communities may have contributed to forest ecosystem functioning, especially the availability of biologically available N, we analysed the correlation between the diversity of diazotrophic communities and the biologically available N and soil carbon (C) (Fig. 2A–C). The Chao1 richness of diazotrophic communities was significantly ($P < 0.005$) correlated with soil NH_4^+ and NO_3^- concentrations, with Pearson correlations of 0.35 and 0.25, respectively (Fig. 2A, B). Interestingly, the Chao1 richness of soil diazotrophs was negatively ($P < 0.005$) correlated with total soil C (Fig. 2C), indicating organic decomposers might be an important contributor of biologically available N when plenty of organic matter is available in the soil. Such significant correlations between microbial diversity and soil N cycling have been observed previously (Philippot *et al*. 2013).
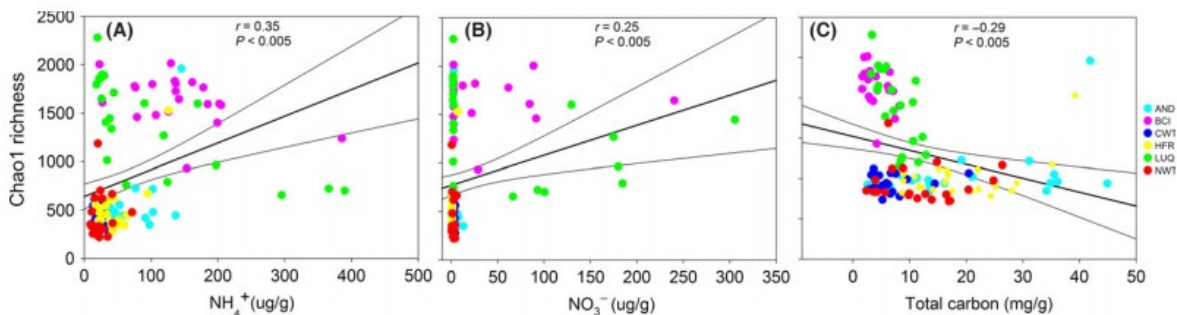

Fig. 2 Correlation between *nifH* OTU richness and soil nitrogen and carbon, including soil NH_4^+ concentrations (A), NO_3^- concentrations (B) and total carbon (C).

Rainforest ecosystems have lower taxa–area relationships

With the experimental design and sampling effort in this study, we were able to analyse the TARs for soil diazotrophic communities in different ecosystems. Here, the *z*-value representing the slope of the taxa–area relationship in a log–log space was calculated for diazotrophic communities in the different ecosystems. Interestingly, rainforests showed significantly ($P = 0.002$) lower *z*-values than that in temperate forest ecosystems (Fig. 3). The two rainforest ecosystems, LUQ and BCI, were observed to have taxa–area relationship *z*-values of 0.06 and 0.061, respectively. The other four temperate forests (AND, CWT, NWT and HRF) were found to have *z*-values of 0.114, 0.099, 0.117 and 0.103, respectively. Such markedly different *z*-values for the soil diazotrophic communities between rainforests and temperate forests indicated different biogeographic patterns of soil diazotrophs in these two environments.
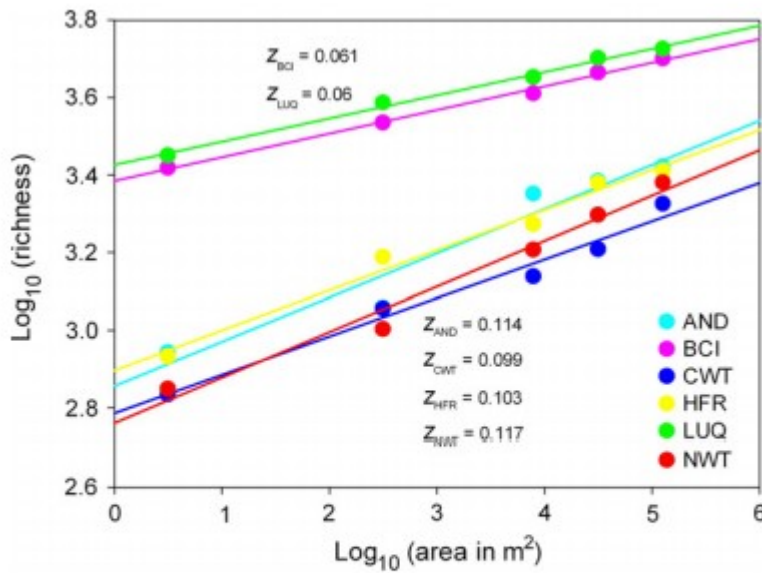
Fig. 3 The TAR (log-transformed) of *nifH* community across the six forest ecosystems. Significantly lower z-values were observed for rainforest ecosystems than temperate forest ecosystems.

The relationships between z-values and individual environmental parameters, including latitude, annual temperature, plant richness, annual precipitation, soil moisture and pH, were also analysed (Fig. 4). Interestingly, the taxa–area relationship z-values were strongly and significantly correlated with latitude ($r = 0.954$, $P = 0.003$) and annual temperature ($r = 0.925$, $P = 0.008$) (Fig. 4A, B), followed by plant richness ($r = -0.82$, $P = 0.046$) and annual precipitation ($r = -0.83$, $P = 0.042$) (Fig. 4C, D). For soil moisture and pH, no significant correlations with z-values were observed (Fig. 4E, F). This suggested that the observed taxa–area relationship for soil diazotrophic communities also followed biogeographic gradient patterns.
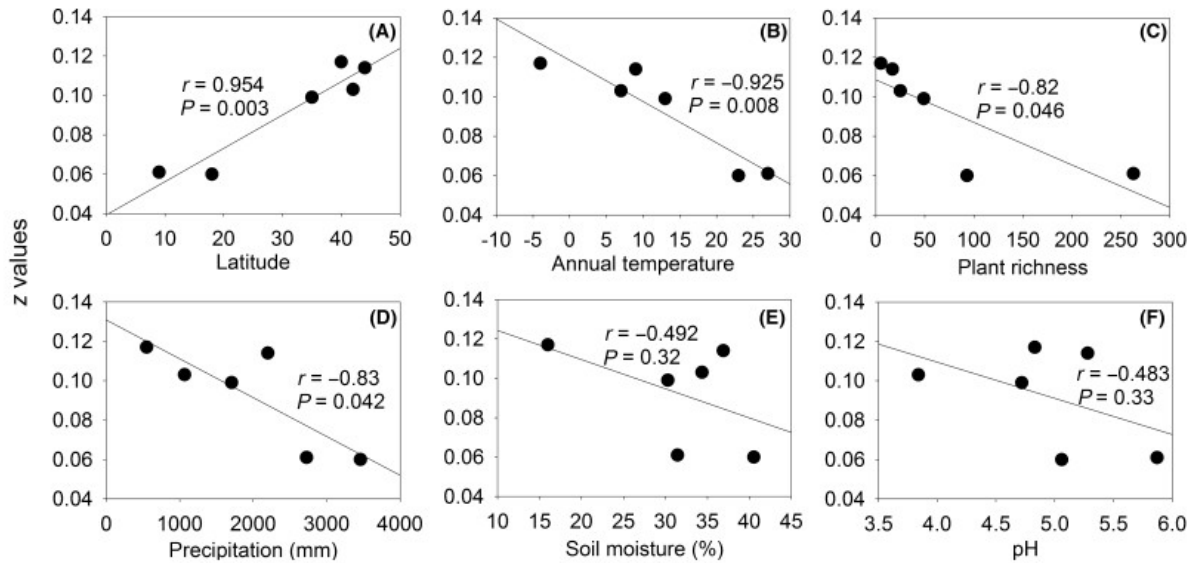
Fig. 4 Correlation between the slopes of TARs and environmental factors, including latitude (A), annual temperature (B), plant species richness (C), annual precipitation (D), soil moisture (E) and pH (F). No significant correlations could be observed between z-values and soil moisture and pH, but latitude, annual temperature, plant richness and precipitation.

Biogeographic diversity gradients of forest diazotrophic communities

To investigate whether the diversity of soil diazotrophic communities was associated with abiotic and biotic properties of different forest ecosystems, correlation analysis was performed between the *nifH* species richness and latitude, temperature, plant richness, precipitation, moisture and pH (Fig. 5). Clear biogeographic gradient patterns of *nifH* species richness could be observed. Among these, latitude was the strongest factor correlated with *nifH* species richness, with Pearson correlation of -0.83 and a P-value < 0.005 (Fig. 5A). Notably, three other factors related to latitude, including annual mean temperature, plant richness and precipitation, were also found to be significantly ($P < 0.005$) correlated with *nifH* species richness, with correlation of 0.76, 0.76 and 0.7, respectively (Fig. 5B–D). Relatively lower correlations were found between *nifH* richness and soil moisture or soil pH, 0.27 and 0.5 ($P < 0.005$), respectively (Fig. 5E, F). These results indicated that *nifH* communities were affected by both abiotic and biotic biogeographic properties of different forest ecosystems, but most strongly by factors related to latitude, such as annual temperature, plant species richness and precipitation. This was also confirmed by canonical correspondence analysis, a multiple regression method, which showed latitude was the strongest factor in shaping the soil diazotrophic community structure, followed by temperature, precipitation and plant species richness (Fig. S5, Supporting information).
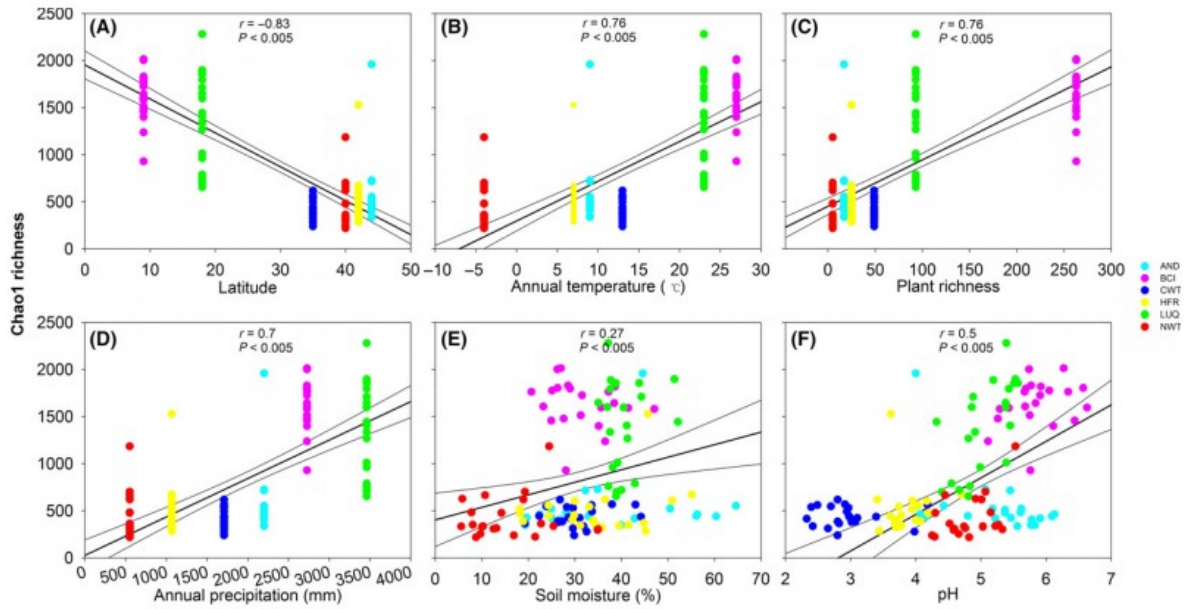
Fig. 5 Correlation between *nifH* OTU richness and environmental factors, including latitude (A), annual temperature (B), plant species richness (C), precipitation (D), moisture (E) and pH (F).

Based on the above information, we present a biogeographic model for forest soil diazotrophic communities (Fig. 6). In this model, as is already commonly known in ecology, latitudinal gradient patterns were observed for several major ecosystem properties, including annual mean temperature, precipitation and plant richness, with correlation value of -0.87, -0.74 and -0.92 ($P < 0.005$). The plant richness was also significantly ($P < 0.005$) correlated with annual temperature, with a correlation value of 0.83. Together, these four factors were the major factors responsible for the biogeographic gradient distribution of soil diazotrophic species richness, with absolute correlation values >0.7. Several other factors, such as soil moisture, pH and total C, were also found to be significant ($P < 0.005$), but not as strongly correlated with the soil diazotrophic species richness. Among these, soil moisture and pH were positively correlated (correlation of 0.27 and 0.5, $P < 0.005$), while total soil C was negatively correlated (correlation of -0.29, $P < 0.005$). Importantly, the species richness of soil diazotrophs was also positively correlated with NH_4^+ and NO_3^- concentrations in soil, suggesting the importance of soil diazotrophs in regulating ecosystem functioning.
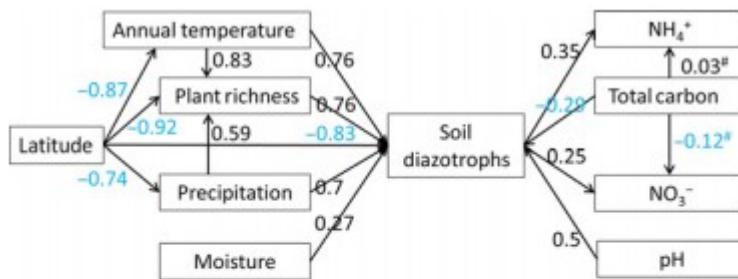
Fig. 6 A biogeographic model for the soil diazotrophic community and the relationship between and within environmental factors. Numbers represent correlation values between environmental factors and *nifH* OTU richness. [#]indicates insignificant correlations.

Discussion

Soil diazotrophic communities are of crucial importance in ecosystem functioning. However, whether they follow similar biogeographic patterns to macroorganisms and other microbial communities has been rarely investigated. In this study, apparent biogeographic gradient patterns of community diversity and taxa–area relationships were observed. The highest diversity of diazotrophic communities in tropical forests and lower diversity of temperate forests generated a traditional latitudinal diversity gradient. The gradient patterns of soil diazotrophic communities were strongly and significantly correlated with latitude, annual temperature, plant richness and precipitation. These results expand our understanding of the biogeographic patterns of soil diazotrophs and provide evidence that soil microbial communities also followed biogeographic diversity gradients as observed in macroorganisms.

We hypothesized that different types of forests may harbour different soil diazotrophic communities. Consistent with our expectation, several interesting phenomena were observed. First, different forest ecosystems harbour distinct soil diazotrophic communities, with different dominant N_2 fixers in particular, indicating a clear biogeographic distribution of soil diazotrophic communities in forest ecosystems. Notably, such distinct community structures were even observed at the class level. Second, all four temperate forests were dominated by a few OTUs with high relative abundances, whereas no OTUs with >4% relative abundance detected in the rainforest ecosystems. The dominance of a few *nifH* OTUs was also previously observed in a grassland ecosystem, where the 15 most abundant OTUs accounted for 67.8% of the total captured *nifH* sequences in that ecosystem (Tu *et al*. 2015). Third, consistent with what has been observed for macrocommunities such as plant and animal communities, rainforests were found to have *nifH* OTU richness values twice that observed for temperate forest ecosystems. All of these observations suggest an obvious

biogeographic distribution of soil diazotrophs across different forest ecosystems. Notably, previous studies also suggested that soil diazotrophic community composition and diversity were associated with location, soil type, organic carbon concentration (Poly *et al*. 2001; Zehr *et al*. 2003; Wang *et al*. 2013; Collavino *et al*. 2014), water temperature (Langlois *et al*. 2008) and nitrogen gradients (Fierer *et al*. 2012). This is striking in that it suggests that microbial subcommunities with specific functions such as soil diazotrophs are also subject to biogeographic changes although they may only occupy a small portion of the soil microbial community.

Our second hypothesis was that soil diazotrophs may also follow ecological principles such as TARs and biogeographic diversity gradients. TAR is one of a few general ecological laws followed by both microbial and macrocommunities (Lawton 1999; Green *et al*. 2004; Horner-Devine *et al*. 2004; Green & Bohannan 2006; Zhou *et al*. 2008; Liang *et al*. 2015), such as plants and animals, and is fundamental to understanding the distribution of biodiversity (Horner-Devine *et al*. 2004). Several studies over the past decade have provided evidence for the existence of a microbial TAR (Green *et al*. 2004; Horner-Devine *et al*. 2004; Green & Bohannan 2006; Peay *et al*. 2007; Zhou *et al*. 2008; Liang *et al*. 2015). The observed z-values varied from as low as 0.01 (salt march β-proteobacteria) (Horner-Devine *et al*. 2004) to as high as 0.23 (soil ectomycorrhizal fungi) (Peay *et al*. 2007) in previous studies. Such differences could be due to different methods used for data analysis (e.g. different thresholds for OTU definition) (Horner-Devine *et al*. 2004), varying ecosystems, as well as divergent habitats for different microorganisms. Notably, different functional and phylogenetic groups harbour different TARs, as was reported in a previous study (Zhou *et al*. 2008). Interestingly, although different z-values were observed for soil diazotrophs in different forest ecosystems, that is ~0.06 for tropical forest ecosystems and ~0.1 for temperate forest ecosystems, the values were generally comparable to a previous observation of *nifH* genes (~0.09) detected by GeoChip in a deciduous forest at the Oak Ridge Reservation (Zhou *et al*. 2008).

However, although much work has been done during the past few years, whether the TARs of microbial communities follow any biogeographic gradient pattern (e.g. latitudinal, temporal, pH and saline) was unclear. Intriguingly, a significant gradient pattern of TAR was observed with tropical rainforest ecosystems having significantly lower z-values than temperate forest ecosystems. Among the environmental factors related to latitude, z-values were strongly and negatively correlated with annual temperature, plant species richness and precipitation. Because higher plant species richness would create more ecological niches for soil diazotrophs (i.e. higher habitat heterogeneity), higher z-values were expected for rainforest ecosystems. The unexpected lower z-values in rainforest ecosystems support the metabolic theory hypothesis (Brown *et al*. 2004) that high temperature

influences microbial community ecological processes, such as competition and dispersal (Connor & McCoy 1979; Peay *et al*. 2007; Báldi 2008; Shen *et al*. 2009). This was also reflected by the fact that in rainforest ecosystems, no *nifH* OTUs contributed >5% relative abundance for the whole *nifH* data set. However, the extent to which dispersal (decreased beta diversity) and habitat heterogeneity (increased beta diversity) contribute to the TAR cannot be determined, especially that no significant differences were found for the within-site beta diversity between rainforest and temperate forest ecosystems.

In addition, the different z-values between rainforest and temperate forest ecosystems also provide evidence that the observed power-law TAR is a generic ecological principle rather than random sampling effects. The species–area relationship theory was introduced in the 1960s and hypothesized that larger areas tend to contain larger number of species and that such relationships follow systematic mathematical relationships (Preston 1960, 1962; Williams 1964). However, for many years, arguments have been made concerning whether the observed power-law species–area relationship is a generic ecological rule or rather simply a result of random sampling effects (Connor & McCoy 1979; Coleman 1981; Coleman *et al*. 1982). This question is especially important in microbial ecology where random sampling plays an important role in underestimating community diversity (Zhou *et al*. 2008, 2011). Theoretically, if random sampling were the dominant factor responsible for the TAR, a similar z-value would be expected for different forest ecosystems when examined by the same technology. In contrast, significantly lower z-values were observed for tropical rainforest ecosystems than for temperate forest ecosystems, confirming the TAR as a generic ecological law.

Another important ecological question that can be addressed in this study is whether microbial communities follow biogeographic diversity gradient patterns similar to those followed by macroorganisms. Clear biogeographic diversity gradient pattern of *nifH* communities could be observed across the six forest ecosystems. Among all environmental variables we examined, latitude, annual mean temperature, plant species richness and precipitation were the factors that most strongly correlated with *nifH* community richness. However, it should be noted that the contribution of other factors such as precipitation and soil pH should not be ruled out, because the gradient patterns we observed for the microbial community could be the result of the overall effect of these environmental parameters.

Although latitudinal gradients in diversity have been regarded as a generic ecological law, especially for plants and animals (Willig *et al*. 2003; Hillebrand 2004; Mittelbach *et al*. 2007), little is known about the responses of microbial communities to gradients (Martiny *et al*. 2006; Fuhrman 2009). Unlike macrocommunities, questions of whether microbial communities follow latitudinal gradients are still debated (Fenchel & Finlay 2004; Fuhrman 2009). Previous studies have suggested that soil microbial

community richness and diversity follow pH gradients, whereas apparent latitudinal gradients were not observed (Fierer & Jackson 2006; Chu *et al*. 2010; Griffiths *et al*. 2011). However, consistent with the current study, latitudinal gradients were observed at a regional scale in a recent study, although pH was the best explanation the shifts in community structure (Shay *et al*. 2015). Such disagreement in microbial diversity gradients could be due to scale issues and/or the large environmental heterogeneity of collected samples. Interestingly, in marine environments at larger scales and less environmental heterogeneity, clear latitudinal gradient patterns could be observed for planktonic bacteria diversity in two recent studies (Pommier *et al*. 2007; Fuhrman *et al*. 2008). Our observations in this study suggest that terrestrial microbial communities, even subgroups like N_2-fixation communities, also followed a similar latitudinal diversity gradient as macroorganisms, with temperature and plant richness as major factors that correlated with microbial diversity. However, it should be noted that temperature and plant richness were significantly correlated with latitude in this study and that the latitudinal diversity gradients observed here could be regarded as an overall representation of many other environmental factors. Notably, our results show that pH also correlated significantly with diazotrophic community diversity, and support previous studies showing pH gradient patterns (Fierer & Jackson 2006; Chu *et al*. 2010; Griffiths *et al*. 2011).

The significant correlation between *nifH* OTU richness and temperature, plant species richness and precipitation observed in this study supported several previously proposed ecological theories. The first is the metabolic theory of ecology. It is expected that higher temperature could lead to higher metabolic rates, which may strongly influence complex ecological processes such as speciation, competition and dispersal of microbial communities (Brown *et al*. 2004; Fuhrman *et al*. 2008; Fuhrman 2009), resulting in higher microbial diversities in tropical rainforest ecosystems. The second is the habitat heterogeneity hypothesis (Bazzaz 1975; Davidowitz & Rosenzweig 1998; Tews *et al*. 2004). Our results suggest that plant richness is positively correlated with microbial community diversity. This could be due to rainforest ecosystems harbouring a much higher plant species richness than temperate forest ecosystems, providing a greater number of diverse ecological niches for microbial communities, and therefore a higher microbial diversity. The third, although not directly evidenced in this study, is the species–energy relationship (Gaston 2000). The higher temperature, precipitation and plant species richness in rainforest ecosystems result in higher primary productivity, which then provides more usable environmental energy for the microbial communities, leading to an increased microbial diversity. These results indicated that the observed latitudinal diversity gradient patterns of the terrestrial microbial communities in forest ecosystems could be a result of multiple complicated ecological processes.

In summary, this study surveyed the biogeography of N_2-fixing communities by analysing *nifH* gene amplicons across six forest ecosystems in North America. This is one of a few studies where biogeographic properties of microbial communities were analysed in terrestrial ecosystems. Our results show that even microbial subcommunities such as soil diazotrophs follow several generic ecological principles. These observations provide a better understanding of the microbial ecology of soil diazotrophic communities, as well as the biogeography of microbial communities in general.

Data accessibility

Raw sequence data were deposited under NCBI BioProject Accession No. PRJNA308872.

References

Ahn SJ, Costa J, Rettig Emanuel J (1996) PicoGreen quantitation of DNA: effective evaluation of samples pre-or post-PCR. *Nucleic Acids Research*, 24, 2623– 2625.

Báldi A (2008) Habitat heterogeneity overrides the species–area relationship. *Journal of Biogeography*, 35, 675– 681.

Bazzaz FA (1975) Plant species diversity in old-field successional ecosystems in southern Illinois. *Ecology*, 56, 485– 488.

Benson DR, Dawson JO (2007) Recent advances in the biogeography and genecology of symbiotic Frankia and its host plants. *Physiologia Plantarum*, 130, 318– 330.

Benson DR, Vanden Heuvel B, Potter D (2004) Actinorhizal symbioses: diversity and biogeography. *Plant Microbiology* (ed M Gillings A Holmes), pp. 97– 127, Scientific Publishers Ltd, Oxford.

Berthrong S, Yeager CM, Gallegos-Graves L *et al*. (2014) Nitrogen fertilization has a stronger effect on soil N-fixing bacterial communities than elevated atmospheric CO_2. *Applied and Environmental Microbiology*, 80, 3103– 3122.

Bremner JM (1965) Inorganic forms of nitrogen. In: *Methods of Soil Analysis. Part 2* (ed CA Black), pp. 1179– 1237. Agron. Monogr. 9. ASA and SSSA, Madison, Wisconsin.

Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. *Ecology*, 85, 1771– 1789.

Caporaso JG, Lauber CL, Walters WA *et al*. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, 6, 1621– 1624.

Chu HY, Fierer N, Lauber CL *et al*. (2010) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology*, 12, 2998– 3006.

Cleveland CC, Townsend AR, Schimel DS *et al*. (1999) Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. *Global Biogeochemical Cycles*, 13, 623– 645.

Coleman BD (1981) On random placement and species-area relations. *Mathematical Biosciences*, 54, 191– 215.

Coleman BD, Mares MA, Willig MR, Hsieh Y-H (1982) Randomness, area, and species richness. *Ecology*, 63, 1121– 1133.

Collavino MM, Tripp HJ, Frank IE *et al*. (2014) nifH pyrosequencing reveals the potential for location-specific soil chemistry to influence N2-fixing community dynamics. *Environmental Microbiology*, 16, 3211– 3223.

Connor EF, McCoy ED (1979) The statistics and biology of the species-area relationship. *American Naturalist*, 113, 791– 833.

Davidowitz G, Rosenzweig ML (1998) The latitudinal gradient of species diversity among North American grasshoppers (Acrididae) within a single habitat: a test of the spatial heterogeneity hypothesis. *Journal of Biogeography*, 25, 553– 560.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996– 998.

Fenchel T, Finlay BJ (2004) The ubiquity of small species: patterns of local and global diversity. *BioScience*, 54, 777– 784.

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

Fierer N, Lauber CL, Ramirez KS *et al*. (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME Journal*, 6, 1007– 1017.

Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science*, 296, 1061– 1063.

Follows MJ, Dutkiewicz S, Grant S, Chisholm SW (2007) Emergent biogeography of microbial communities in a model ocean. *Science*, 315, 1843– 1846.

Franche C, Lindström K, Elmerich C (2009) Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant and Soil*, 321, 35– 59.

Fuhrman JA (2009) Microbial community structure and its functional implications. *Nature*, 459, 193– 199.

Fuhrman JA, Steele JA, Hewson I *et al*. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences of the USA*, 105, 7774– 7778.

Galloway JN, Dentener FJ, Capone DG *et al*. (2004) Nitrogen cycles: past, present, and future. *Biogeochemistry*, 70, 153– 226.

Gaston KJ (2000) Global patterns in biodiversity. *Nature*, 405, 220– 227.

Gentry A (1982) Patterns of neotropical plant species diversity. *Evolutionary Biology*, 15, 1- 84.

Ghiglione J-F, Galand PE, Pommier T *et al*. (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences of the USA*, 109, 17633– 17638.

Green J, Bohannan BJ (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution*, 21, 501– 507.

Green JL, Holmes AJ, Westoby M *et al*. (2004) Spatial scaling of microbial eukaryote diversity. *Nature*, 432, 747– 750.

Griffiths RI, Thomson BC, James P *et al*. (2011) The bacterial biogeography of British soils. *Environmental Microbiology*, 13, 1642– 1654.

Groszkopf T, Mohr W, Baustian T *et al*. (2012) Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature*, 488, 361– 364.

Hillebrand H (2004) On the generality of the latitudinal diversity gradient. *American Naturalist*, 163, 192– 211.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ (2004) A taxa–area relationship for bacteria. *Nature*, 432, 750– 753.

Hsu S-F, Buckley DH (2009) Evidence for the functional significance of diazotroph community structure in soil. *ISME Journal*, 3, 124– 136.

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research*, 17, 377– 386.

Izquierdo J, Nüsslein K (2006) Distribution of extensive nifH gene diversity across physical soil microenvironments. *Microbial Ecology*, 51, 441– 452.

Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the USA*, 102, 2567– 2572.

Ladau J, Sharpton TJ, Finucane MM *et al*. (2013) Global marine bacterial diversity peaks at high latitudes in winter. *ISME Journal*, 7, 1669– 1677.

Langlois RJ, Hümmer D, LaRoche J (2008) Abundances and distributions of the dominant nifH phylotypes in the Northern Atlantic Ocean. *Applied and Environmental Microbiology*, 74, 1922– 1931.

Lawton JH (1999) Are there general laws in ecology? *Oikos*, 84, 177– 192.

Liang Y, Wu L, Clark IM *et al*. (2015) Over 150 years of long-term fertilization alters spatial scaling of microbial biodiversity. *mBio*, 6, e00240– 15.

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957– 2963.

Martiny JBH, Bohannan BJ, Brown JH *et al*. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102– 112.

Ministry of Agriculture FaFM-ADASA (1986) *The Analysis of Agricultural Materials: A Manual of the Analytical Methods Used by the Agricultural Development and Advisory Service*, 3rd edn. HMSO, London.

Mittelbach GG, Schemske DW, Cornell HV *et al*. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, 10, 315– 331.

Mohamed NM, Colman AS, Tal Y, Hill RT (2008) Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environmental Microbiology*, 10, 2910– 2921.

Moisander PH, Shiue L, Steward GF *et al*. (2006) Application of a nifH oligonucleotide microarray for profiling diversity of N2-fixing microorganisms in marine microbial mats. *Environmental Microbiology*, 8, 1721– 1735.

Oksanen J, Kindt R, Legendre P *et al*. (2007) The vegan package. Community ecology package, 631– 637.

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. *Ecology Letters*, 10, 470– 480.

Philippot L, Spor A, Henault C *et al*. (2013) Loss in microbial diversity affects nitrogen cycling in soil. *ISME Journal*, 7, 1609– 1619.

Poly F, Ranjard L, Nazaret S, Gourbière F, Monrozier LJ (2001) Comparison of nifH gene pools in soils and soil microenvironments with contrasting properties. *Applied and Environmental Microbiology*, 67, 2255– 2262.

Pommier T, CanbÄCk B, Riemann L *et al*. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, 16, 867– 880.

Preston FW (1960) Time and space and the variation of species. *Ecology*, 41, 611– 627.

Preston FW (1962) The canonical distribution of commonness and rarity: Part I. *Ecology*, 43, 185– 215.

Schloss PD, Westcott SL, Ryabin T *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing

and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537– 7541.

Shay P-E, Winder RS, Trofymow J (2015) Nutrient-cycling microbes in coastal Douglas-fir forests: regional-scale correlation between communities, in situ climate, and other factors. *Frontiers in Microbiology*, 6, 1097.

Shen G, Yu M, Hu X-S *et al*. (2009) Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology*, 90, 3033– 3041.

Tedersoo L, Bahram M, Põlme S *et al*. (2014) Global diversity and geography of soil fungi. *Science*, 346, 1256688.

Tews J, Brose U, Grimm V *et al*. (2004) Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography*, 31, 79– 92.

Tu Q, Zhou X, He Z *et al*. (2015) The diversity and co-occurrence patterns of N2-fixing communities in a CO_2-enriched grassland ecosystem. *Microbial Ecology*, 71, 604– 615.

Vitousek PM, Howarth RW (1991) Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry*, 13, 87– 115.

Wang Q, Quensen JF, Fish JA *et al*. (2013) Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio*, 4, e00592– e00513.

Williams C (1964) *Patterns in the Balance of Nature*. Academic Press, London, UK.

Willig MR, Kaufman DM, Stevens RD (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, 34, 273– 309.

Wu L, Wen C, Qin Y *et al*. (2015) Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiology*, 15, 125.

Young J (2005) The phylogeny and evolution of nitrogenases. *Genomes and Genomics of Nitrogen-fixing Organisms* (eds R Palacios, WE Newton), pp. 221– 241. Springer Dordrecht, the Netherlands.

Zehr JP (2011) Nitrogen fixation by marine cyanobacteria. *Trends in Microbiology*, 19, 162– 173.

Zehr JP, Jenkins BD, Short SM, Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology*, 5, 539– 554.

Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 62, 316– 322.

Zhou J, Kang S, Schadt CW, Garten CT (2008) Spatial scaling of functional gene diversity across various microbial taxa. *Proceedings of the National Academy of Sciences of the USA*, 105, 7768– 7773.

Zhou J, Wu L, Deng Y *et al*. (2011) Reproducibility and quantitation of amplicon sequencing-based detection. *ISME Journal*, 5, 1303– 1313.