**Meta-analysis of diazotrophic signatures across terrestrial ecosystems at the continental scale**

Chen Zhu 1, Ville-Petri Friman 2, Ling Li ­1, Qicheng Xu 1, Junjie Guo 1, Shiwei Guo 1, Qirong Shen 1, Ning Ling 1, \*

**Affiliations:**

1, *Jiangsu Provincial Key Lab for Solid Organic Waste Utilization, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, Nanjing Agricultural University, Nanjing, 210095, China*

2, *Department of Biology, University of York, Wentworth Way, York, YO10 5DD, UK*

**\*Corresponding author**:

Ning Ling: Tel: +86 25 84396853; E-mail: [nling@njau.edu.cn](mailto:nling@njau.edu.cn)

**Originality-Significance Statement**

Biological nitrogen fixation (BNF) plays an extremely important role for the cycling of nitrogen across earth's ecosystems. However, our knowledge regarding the distribution, and diversity of free-living nitrogen-fixing prokaryotes, diazotrophs, across terrestrial ecosystems is limited. In this study, we use meta-analysis approach to create a comprehensive diazotroph-associated *nifH* gene reference database based on 1,988 data sets derived from 39 independent studies. By using multiple computational approaches, we identify key diazotroph indicator species, including *Azospirillum zeae*, *Skermanella aerolata* and four *Bradyrhizobium* species, that are associated to specific terrestrial ecosystems. Moreover, our analyses further identify potential environmental threshold values that could determine the diazotroph abundances and distribution in soil ecosystems regarding soil pH, nitrogen, organic carbon, C:N ratio and annual mean precipitation and temperature. This study represents an important step for better understanding the diversity and distribution of diazotrophs across terrestrial ecosystems globally.

**Summary**

Biological nitrogen fixation (BNF) performed by diazotrophs forms a cornerstone of Earth’s terrestrial ecosystem productivity. However, the composition, diversity and distribution of soil diazotrophs is poorly understood across different soil ecosystems. Furthermore, the biological potential of the key diazotroph species in relation to key environmental parameters is unknown. To address this, we used meta-analysis approach to merge together 39 independent diazotroph amplicon sequencing (*nifH* gene) datasets consisting of 1,988 independent soil samples. We then employed multiple statistical analyses and machine-learning approaches to compare diazotroph community differences and indicator species between terrestrial ecosystems on a global scale. The distribution, composition and structure of diazotroph communities varied across seven different terrestrial ecosystems, with community composition exhibiting an especially clear effect. The *Cyanobacteria* were the most abundant taxa in crust ecosystems (accounting for ~45% of diazotrophs), while other terrestrial ecosystems were dominated by *Proteobacteria*,including *Alpha*-, *Beta*- and *Gamma*-*Proteobacteria* (accounting for ~70% of diazotrophs). Farmland ecosystems harboured the highest and crust ecosystems the lowest alpha and phylogenetic diversities. *Azospirillum zeae*, *Skermanella aerolata* and four *Bradyrhizobium* species were identified as key indicator species of potential diazotroph activity. Overall, diazotroph abundances and distribution were affected by multiple environmental parameters, including soil pH, nitrogen, organic carbon, C:N ratio and annual mean precipitation and temperature. Together, our findings suggest that based on the relative abundance and diversity of *nifH* marker gene, diazotrophs have adapted to a range of environmental niches globally.

**Keywords:** diazotrophs; terrestrial ecosystems; continental scale; soil samples; community composition; keystone diazotroph species; meta-analysis

**Introduction**

As the principal building block of nucleic acids and proteins, nitrogen (N) is a fundamental element of life, often limiting the primary productivity of ecosystems (LeBauer and Treseder 2008; Vitousek and Howarth 1991; Xu et al. 2015). The main form of N in nature is dinitrogen (N2), which cannot be directly utilized by most organisms except for a subset of microbes (Cheng 2008; Reed et al. 2011). Biological N fixation (BNF) thus plays a key role in converting the abundant but inert N2 molecule into ammonia and into biologically available N (Galloway et al. 2004; Vitousek et al. 2013). BNF is thus crucial for the global N cycle and regarded as the cornerstone of biotic life on earth alongside with photosynthesis (Cheng 2008).

The energetically demanding, and oxygen-sensitive reaction of BNF is catalysed by the metalloenzyme nitrogenase, which is found in free-living (*i.e.*, diazotrophs) and plant symbiotic or mutualistic prokaryotes (Gaby and Buckley 2011). Although it is generally accepted that most nitrogen fixation in terrestrial ecosystems is carried out by symbiotic rhizobia (Peoples and Craswell 1992; Cleveland et al. 1999; Herridge et al. 2008), free-living nitrogen-fixing bacteria may be vital contributors to BNF in environments where legumes are absent (Vitousek et al. 2013). Free-living N2 fixation also occurs in relatively less studied ecosystems (Reed et al. 2013; Matson et al. 2015), such as deep soils and canopy soils, which may lead to an underestimation of their global significance. Compared with symbiotic N fixation, the rate of non-symbiotic N fixation (free-living N fixation) is thought to be much lower (Son 2001). Compared to symbiotic diazotrophs, non-symbiotic diazotrophic species are more widely distributed in space and time, especially in non-agricultural systems where nitrogen is often limited and where they can serve as the primary source of N fixation (Rosen and Allan 2007). However, in many tropical forests, particularly those lacking many symbiotic N-fixing plants, non-symbiotic N fixation is extremely important allowing plants to meet their high N demand and balance the loss of N from the soil (Chalk et al. 2017). Although it is generally accepted that rhizobia only fix nitrogen in the nodules of plants (Lee et al. 2008; Dreyfus et al. 1988), some members of *Azorhizobium* and *Bradyrhizobium* have been shown to fix N2 in both the symbiotic and free-living states (Dreyfus et al. 1983; Alazard 1990). For example, a non-symbiotic strain *Bradyrhizobium* sp. AT1 has been reported to fix N2 with a rate comparable to the photosynthetic strain *Bradyrhizobium* sp. ORS278 in free-living state (Terakado-Tonooka and Ohwaki 2013). Another example comes from *Bradyrhizobium* sp. DOA9, which harbors a *nif* cluster on its chromosome and another on a symbiotic plasmid. Interestingly, chromosomal *nif* cluster of *Bradyrhizobium* sp. DOA9 is involved in the fixation of N2 in the free-living state. (Wongdee et al. 2018). As such, the investigation of diazotrophs in soils is of great importance for understanding the productivity of natural terrestrial ecosystems.

Thus far, there have been surveys on the environmental prevalence and diversity of diazotrophic prokaryotes in marine (Langlois et al. 2005; Turk et al. 2011) and estuarine sediments (Affourtit et al. 2001), terrestrial geothermal springs (Hall et al. 2008; Hamilton et al. 2011) and terrestrial soils (Hamelin et al. 2002; Wang et al. 2017). The most dominant drivers for diazotrophic community structure are pH (Levy-Booth et al. 2014; Tu et al. 2016), organic matter (Gupta et al. 2014; Wakelin et al. 2010) and the carbon to N ratio of the soil (Wang et al. 2017). Decades of field studies and combinations of molecular biology, isotope-based analyses, geochemical modelling and next-generation sequencing approaches have broadened the understanding of ecology of diazotrophic prokaryotes. For example, Zehr and Capone (2020) reported that diazotrophic bacteria are often adapted to extreme environmental conditions in marine environments such as to a high pressure. These findings, in combination with diagnostic distributions of nutrients and their isotopes, as well as measured and modelled biogeographic patterns, have revolutionized our understanding of marine N2 fixation and its role in the global N cycle. Several studies have explored the role of diazotrophs in different terrestrial ecosystems. For example, long-term fertilization (especially N fertilization) may select against N fixation, create a bias towards specific groups of N fixers and drastically reduce N-fixing diazotroph abundances and diversity in farmland ecosystems (Fan et al. 2019). In karst grassland ecosystems, the diversity and composition of the diazotroph community were found to be more sensitive to phosphorus addition than to N addition, with phosphorus availability appearing to regulate N2 fixation by increasing diazotrophic diversity (Xiao et al. 2020). BNF is further shaped by other environmental factors, and for example the abundance, diversity, and community composition of soil diazotrophs has been shown to follow the mean annual precipitation in Tibetan grassland soils (Che et al. 2018). Moreover, various diazotrophic taxa exhibit deterministic abundance-elevation relationships, contributing to composition at both low and high elevations within forest ecosystems (Wang et al. 2019), while stochastic processes have been found to be more significant in determining diazotroph distribution at intermediate elevations (Wang et al. 2019). Moreover, diazotrophic communities have been found to assemble along gradients of environmental factors, being particularly impacted by cation-exchange capacity of soils (Lee et al. 2019). Collectively, these studies suggest that the distribution and diversity of diazotrophs varies across ecosystems, which could be determined by differences in environmental conditions. While a recent meta-analysis on global patterns of BNF under nutrient enrichment exists (Zhang et al. 2018), a comprehensive sequence-based study systematically linking diazotroph distribution and environmental factors across terrestrial ecosystems is missing. To fill this knowledge gap, we built a diazotroph-associated *nifH* gene reference database encompassing 2828 high quality-filtered sequences of a minimum of 261 bp length, based on 1,988 data sets derived from 39 independent studies. We then used meta-analysis and multiple computational approaches to analyse diazotroph diversity and community composition based on *nifH* gene presence and diversity across soil ecosystems and to identify key diazotroph indicator species. Please note, while the *nifH* gene presence does not necessary indicate active BNF, we use it as an indicator of potential diazotroph activity throughout the manuscript.

The specific objectives of this study were to (1) to improve our understanding about the diversity and distribution of diazotrophs among various terrestrial ecosystems; (2) to determine keystone diazotrophic species associated with specific ecosystems, and (3) to investigate correlations between abiotic environmental characteristics and diazotroph species distribution and community composition. By combining multiple statistical analyses and machine-learning approaches, this meta-analysis study provides a comprehensive overview of diazotrophic communities across terrestrial ecosystems and pinpoints potential environmental thresholds that might explain their distribution.

**Methods**

**Collecting data for meta-analysis of diazotrophs**

We searched for articles on N-fixing prokaryotes and corresponding sequencing data published before July 2020 (Circa around ~2000 to 2020) using Google Scholar (http://scholar.google.com/), the China National Knowledge Infrastructure (http://www.cnki.net/) and the National Centre for Biotechnology Information (NCBI) SRA database (https://www.ncbi.nlm.nih.gov/). The keywords used for the article and data selection were: (a) “*nifH*” or “diazotrophs” or “nitrogen fixation” or “diazotrophic” and (b) “terrestrial” or “soil” or “land”.

The articles selected for this meta-analysis were required to contain high-throughput sequencing data of the *nifH* gene (based on PCR amplicon sequencing data) and basic metadata information of the samples (at least including the information on the ecosystem types of samples). Some missing data was acquired by directly contacting the authors. After initial article collection, we used the following rules to exclude some of the articles and sequencing data: (a) absence or inaccuracy of accession numbers of high-throughput sequencing data (*e.g.*, providing the wrong accession number or sequencing data from multiple samples were combined and could not be separated); and (b) high-throughput sequencing data from [mutualistic](javascript:;) [symbiosis](javascript:;) *(e.g.*, moss or plant roots). When the same high-throughput sequencing data were used in different studies, it was included only once in this meta-analysis. Based on these criteria, 39 studies comprising of 1,988 high-throughput sequencing data sets were selected for the final meta-analysis and the basic information and accession numbers of samples are provided in supplementary materials (Appendix 1).

**High-throughput sequencing data processing and establishment of the *nifH* gene database**

The raw high-throughput sequencing data was downloaded based on the NCBI Accession Numbers to obtain FASTQ files for each individual data set. The paired-end sequences of each individual study were merged with the *fastq\_mergepairs* command and quality filtered with *fastq\_filter* command in USEARCH software (Edgar 2013), trimming bases below a quality score of 20 and shorter than 240 base pairs. Next, *fastx\_uniques* and *Unosie3* commands were implemented to remove redundant sequences and chimeras. Denoising was then performed to obtain single nucleotide fragments and the remaining sequences were further converted to amino acid sequences using FrameBot (version 1.2) (https://github.com/rdpstaff/Framebot). Sequences encoding proteins that did not match *nifH* protein sequences, or that contained termination codons, were discarded (Wang et al. 2013).

To compare studies that used different *nifH* gene regions, we adopted a closed-reference workflow, which is a database-dependent approach employing a predefined set of reference sequences with known taxonomy. Briefly, *nifH* sequences were clustered into species and assigned to taxonomy based on *nifH* sequences. The *closed\_ref* command (sequence identity ≥ 97%) in USEARCH was performed to map the obtained fragments to the *nifH* reference database (http://www.drive5.com/usearch/manual/sintax\_downloads.html), whereas fragments that could not be mapped were assigned to unknown sequences.

Most included studies aligned the sequences against the *nifH* reference database of FunGene repository, based on the Ribosomal Database Project (RDP) (http://fungene.cme.msu.edu/). However, *nifH* referenced sequences in FunGene repository included redundant sequences and most database sequences lacked the explicit taxonomic information. To ensure that as much data as possible was included in the meta-analysis, an improved *nifH* reference database was constructed by removing redundant sequences and including additional annotations. The new database was established through following steps: (a) all *nifH* sequences (protein sequences) from the FunGene repository of RDP database were downloaded; (b) the structural domain of the *nifH* proteinfrom the FunGene repository (http://fungene.cme.msu.edu/hmm\_detail.spr?hmm\_id=328) was retrieved and (c) HMMER (version 3.3; http://hmmer.org/download.html) was implemented in order to compare the *nifH* structural domains with the protein database ([UniProt](https://www.uniprot.org/help/about) Database; https://www.uniprot.org/). Protein sequences related to *nifH* were selected through threshold screening (E-value ≤ 1e-5), and matched with NCBI annotation information based on the query number of protein sequences; (d) [Zehr’s laboratory](https://www.jzehrlab.com/) provided a portion of the full-length *nifH* sequences obtained from cultivated diazotrophs and we acquired the unabridged database from the website (https://wwwzehr.pmc.ucsc.edu/nifH\_Database\_Public/); (e) The three *nifH* databases were then combined and redundant and artificial sequences were removed while preserving as many sequences as possible. The new improved and integrated database that assigned and identified *nifH* clusters based on protein homology was deposited and made publicly available at Github (https://github.com/ChenZhu-BIOAnalysis/nifH\_meta-Data-Analysis).

The species table was constructed assigning sequences taxonomically based on the new *nifH* protein-based database using the *otutab* command in USEARCH. The species mapping was conducted for of each data individually in Usearch, and tables were directly imported to R software (version 3.6.1) (Team 2013) and merged into one integrated table for further analyses. The workflow of the processing sequencing data, including the codes and parameter settings was also deposited at Github (https://github.com/ChenZhu-BIOAnalysis/nifH\_meta-Data-Analysis).

**Environmental metadata and bioinformatics analysis**

For each study included in our dataset, a wide range of environmental metadata was collected, including information on study site latitude, longitude, elevation, mean annual temperature (MAT), mean annual precipitation (MAP), pH, moisture, soil organic carbon (SOC), total nitrogen (TN), carbon nitrogen ratio (C: N), total phosphorus (TP), total potassium (TK), available phosphorus (AP), available potassium (AK), ammonium nitrogen (NH4+-N) and nitrate nitrogen (NO3--N). We further classified all samples according to the ecosystems where they were collected from, resulting in seven terrestrial ecosystems (crust, farmland, forest, grassland, mine, riparian and tundra) which were treated as categorical variables in subsequent data analyses. The distribution of sampling points was visualized using a global map using the R packages “*sp*”, “*maptools*” and “*maps*”. Kruskal-Wallis rank sum tests were used to assess the differences in a set of diversity measures (Chao1 index, Shannon index and phylogenetic diversity) among the seven ecosystems (Hollander et al. 2013) in R and were visualized using “*ggplot2*”. PERMANOVA and principal coordinates analysis (PCoA) analyses were used to statistically compare the composition of diazotroph communities among different ecosystems using the “*vegan*” package. The normalized stochasticity ratio (NST) was carried out using the “*tNST*” package (Ning et al. 2019) in R to calculate both stochastic and deterministic ecological processes.

**Detection of diazotrophic keystone indicator species**

Keystone diazotroph species were identified using three different approaches: Indicator Value Analysis, Random Forest (RF) and Threshold Indicator Taxa Analysis (TITAN). These three methods were chosen because they use different algorithms and can together provide a more reliable and robust analysis for identifying potential indicator species.

Indicator Value Analysis (Dufrêne and Legendre 1997) is a widely implemented method to find indicator species and species assemblages that characterize groups of all sites. This index is maximized when all individuals of a species are found in a single group of all sites, while species-enriched indicators occur when the species occur across all sites of a given group.

Random Forest (RF) (Liaw and Wiener 2002) approach expands on the previous method by also quantifying the relative importance of indicator species. We used RF models in R using default parameters with the relative abundances of the diazotroph species as explanatory variables. Cross-validation was performed by the “*rfcv*” function for selecting appropriate features, and the varImpPlot function was used to illustrate the importance of diazotroph taxa features. The indicator species of diazotrophs were calculated by the “*labdsv*” package (Dufrêne and Legendre 1997) in R. These indicators generated by RF model were also referred to as “predicted indicators”.

Finally, Threshold Indicator Taxa Analysis (TITAN) was used to detect changes in taxa distributions along environmental gradients in space or time, and to assess synchrony among taxa change points as evidence for community change thresholds (Baker and King 2010). Based on the data between diazotrophic community and environmental variables, the TITAN test was used to analyze taxon-specific contributions to community change along environmental gradients and to infer environmental change points that impacted the diazotroph communities. Only complete datasets that included data on following environmental parameters were included in the TITAN analyses: soil pH, nitrogen, organic carbon, C:N ratio and annual mean precipitation and temperature. The environmental gradients were divided into two groups: z- which corresponds to negative classifications and responds negatively to an increase in the variable, and z+ which are positive classifications and respond positively to an increase along the gradient (Baker and King 2010; King and Baker 2010; Monk et al. 2018; Nguyen et al. 2018). For each taxon, TITAN determines an optimal point of change as the value that maximizes the association of taxa within both z groups. When this point passes from low to high values, the abundance and frequency of occurrence in group z- will decrease, while group z+ will increase. To determine the accuracy of the change in point values, a 500-repeat bootstrap estimation was implemented. This allows for the generation of two groups (purity and confidence) that evaluate the likelihood of response for each taxon. Purity is defined as the proportion in response to a direction (increase or decrease) when it passes the point of inflection that matches the observed response. Pure indicators are assigned in the same response direction. Confidence is estimated by the proportion of change points that consistently result in the significant grouping of a given taxon. For this study, purity and confidence were considered with ≥ 90 values, and these indicators were considered as environmental threshold indicators. Graphs were generated with the *“ggplot2*” package in R (Team 2013; Wickham 2016) and Threshold Indicator Taxa Analysis was conducted using the “*TITAN2*” package (Baker et al. 2015).

The threshold of a single parameter often cannot reflect the overall basis of the response of the diazotroph community to the environment. As such, environmental thresholds of multiple biological parameters can be analysed concurrently to increase the accuracy of the results. In combination with the above three methods, we identified comprehensive indicator species that were determined as significant using all three analysis methods.

**Results**

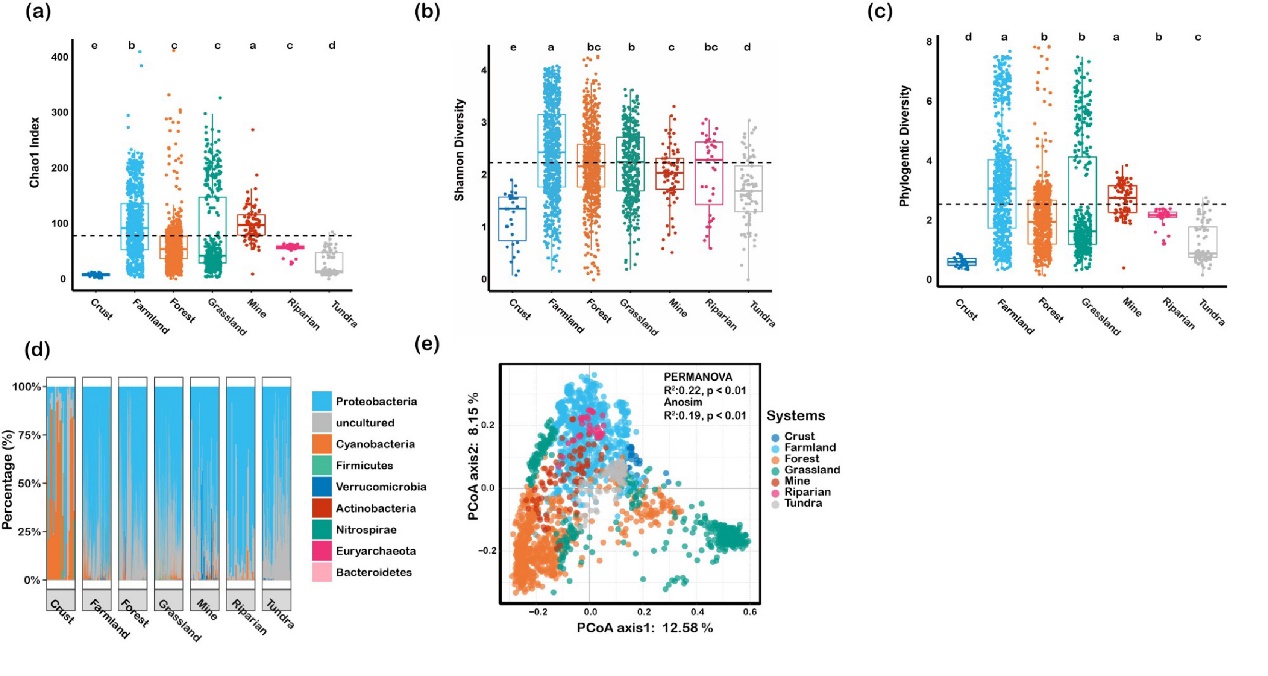
**Global evaluation of diazotroph community composition and diversity across terrestrial ecosystems**

Among the 1,988 datasets covering terrestrial ecosystems across the world, we obtained 30 samples from crust ecosystems (1.5%), 719 samples from farmland ecosystems (36.2%), 654 samples from forest ecosystems (32.9%), 386 samples from grassland ecosystems (19.4%), 78 samples from mine ecosystems (3.9%), 36 samples from riparian ecosystems (1.8%) and 85 samples from tundra ecosystems (4.3%) (Figure S1). The final data set consisted of 48,479,257 high-quality diazotroph *nifH* sequences, which clustered into 1,392 operational taxonomic units (OTUs). Of these sequences, 75.4% (36,562,949 reads) could be assigned taxonomically, while 24.6% of the sequences (11,916,259 reads) belonged to bacteria without taxonomic information. Only a very few sequences belonged to Archaea (49 reads), which could have been due to their low abundance in samples, poor representation of archaeal sequences in databases or primer bias in the original studies. Overall, obtained sequences belonged to eight phyla (*Proteobacteria*, *Cyanobacteria*, *Firmicutes*, *Verrucomircrobia*, *Actinobacteria*, *Nitrospirae*, *Bacteroidetes* and *Euryarchaeota*). *Proteobacteria* was the most abundant phylum in all ecosystems (> 68%), while *Cyanobacteria* was the dominant phylum in the crust ecosystems, accounting for approximately 45% of the total relative abundances (Figure 1d).

To compare the differences in diazotrophic community diversity among different terrestrial ecosystems in more detail, both α- and β-diversity indexes were calculated at the species level (Figure 1). The Chao 1 index, Shannon diversity and phylogenetic diversity of diazotrophs were significantly different among the seven ecosystems (*P* < 0.05) (Figures 1a, 1b and 1c). The Chao1 index ranged from 8.5 to 100.9 and was the highest in the mine ecosystems and the lowest in the crust ecosystems. Relatively high Chao1 values were also observed in the farmland ecosystems (mean = 100.3; Figure 1a). Shannon diversity index values ranged from 1.17 to 2.45 and were the highest in farmland ecosystems and the lowest in crust ecosystems. Slight differences between the forest, grassland, mine and riparian ecosystems were also identified (Figure 1b). The phylogenetic diversity index values ranged from 0.54 to 3.21 with the highest values observed in farmland ecosystems and the lowest values in crust ecosystems. The forest and grassland ecosystems also exhibited high phylogenetic diversity values (Forest: 2.03; Grassland: 2.51), which did not however significantly differ from other ecosystems (Figure 1c).

The variation in diazotrophic community structure was analysed using PCoA method (Figure 1e). The results showed that diazotroph community structure was significantly separated between the seven ecosystems (PERMANOVA: R2=0.22, *P* < 0.001;Anosim: R = 0.19, *P* < 0.001), with the first two principal components explaining 12.58% and 8.15% of the total variation, respectively. Among the three ecosystems with the largest number of samples (farmland, grassland and forest ecosystems), only a very small overlap was observed, indicative of large differences in diazotroph community composition (PERMANOVA: R2=0.13, *P* < 0.001). Much higher overlap was observed between four other ecosystems (crust, tundra, mine and riparian), suggesting that these diazotrophs communities were relatively more similar while still statistically different from each other (PERMANOVA: R2=0.02, *P* = 0.01).

Together these findings suggest that crust ecosystem diazotroph communities were unique, containing a high proportion of *Cyanobacteria*, while other ecosystems were generally more diverse and dominated by *Proteobacteria.* Unexpectedly, farmland soil ecosystems harboured the highest diazotroph diversity and differed from grassland and forest ecosystems regarding diazotroph community composition.

****

**Figure 1. Differences in diazotroph community diversity and composition across soil ecosystems.** Panels (a-c) show the α-diversity indexes of diazotroph communities from seven terrestrial ecosystems (including Chao1 index (a), Shannon diversity (b) and phylogenetic diversity (c). Panel (d) shows phylum-level composition of diazotroph communities. Panel (e) shows Principal Coordinates Analysis (PCoA) of diazotroph community composition based on species matrices from seven terrestrial ecosystems. One-way PERMANOVA was used to analyses the effects of ecosystem-types on community structure of diazotroph communities.

**Predicted species-enriched indicators of diazotrophs in terrestrial ecosystems**

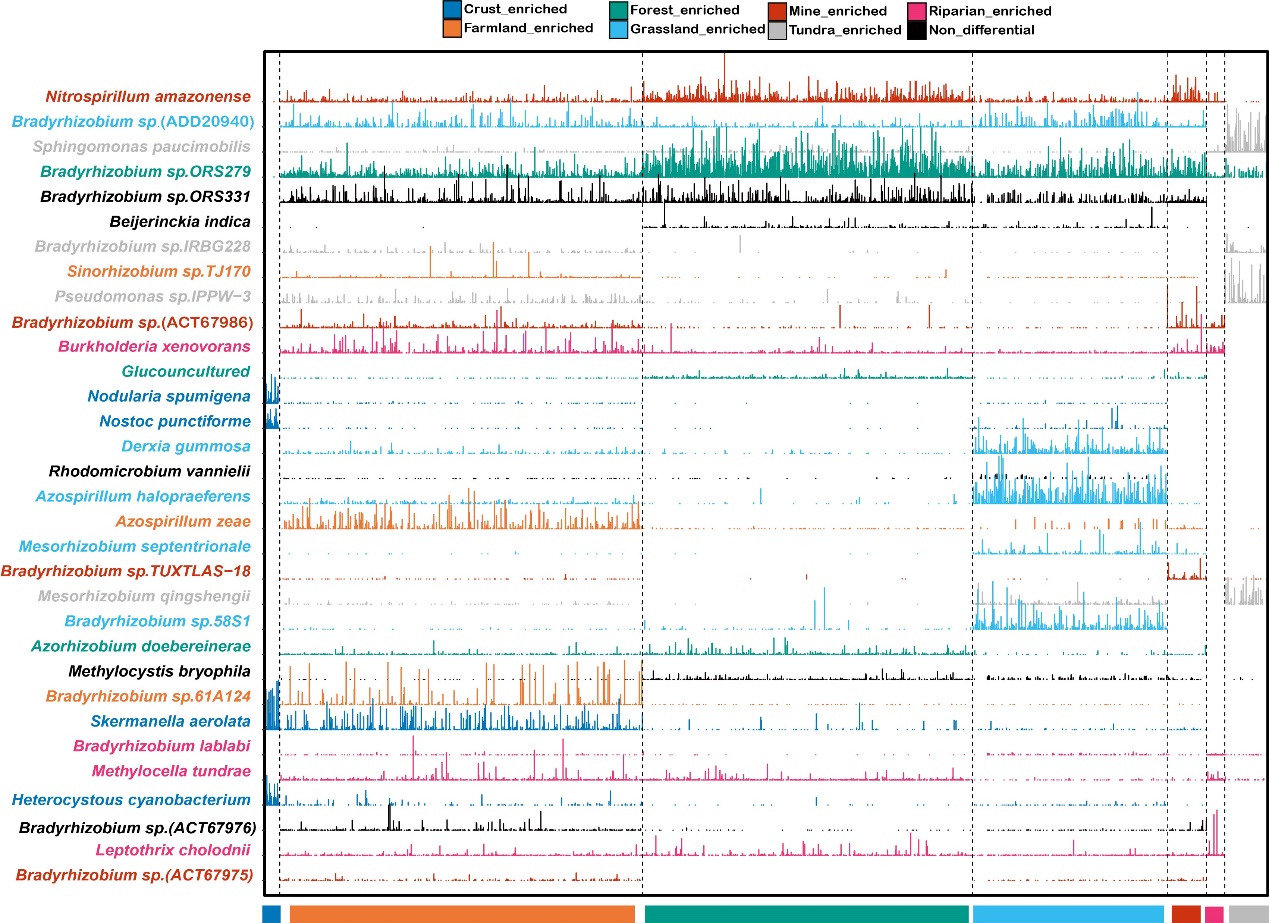
Keystone diazotroph species were identified using Indicator Value Analysis and Random Forest (RF) models. Based on the Indicator Value Analysis, we identified a total of 83 indicator species across the seven ecological system (Appendix. 5). Crust ecosystem was associated with 5 indicator species, grassland ecosystem with 12 indicator species, farmland ecosystems with 10 indicator species, the forest ecosystem with 7 indicator species, the tundra ecosystem with 10 indicator species, the mine ecosystem with 16 indicator species, and the riparian ecosystem with 23 indicator species (Appendix. 5).

RF model and cross-valuation were also used to identify and predict the potential indicator species across seven ecosystems. This analysis resulted in identification of 32 species with prediction accuracies greater than 99% (Figure 2a and 2c). All the predicted indicator species were associated with four phyla, including *Alphaproteobacteria* (25 predicted indicators), *Betaproteobacteria* (3 predicted indicators), *Gammaproteobacteria* (3 predicted indicators) and *Cyanobacteria* (1 predicted indicator). Five of the ten most important predicted indicators belonged to *Bradyrhizobium* (Figure 2b), suggesting that it could be a critical indicator genus for the diazotroph communities across various terrestrial ecosystems.



**Figure 2. Diazotroph indicator species analysis based on Random Forest (RF) model across different soil ecosystems**. A total of 32 bacterial species were identified by applying random-forest classification of the relative abundance of the diazotrophs. Panels (a-b) show predicted indicator species that are ranked in descending order of importance following the accuracy of the model (70% of data used for RF model). (c) The verification of RF model results (30% of remaining data) showing correctly (shown on given ecosystem colors) and incorrectly (black color) predicted indicator diazotrophs in seven ecosystems. In (c), each square corresponds to an individual sampling location.

To distinguish which indicator species had a relatively stronger association with corresponding ecosystem, we combined Indicator Species Analysis with RF analysis. The combined results from the RF model and Indicator Species Analysis revealed that nearly half of all predicted indicator species belonged to *Bradyrhizobium* genus and these indicators were observed in all terrestrial ecosystems except for the crust ecosystems. Instead, four unique indicator species were found in the crust ecosystems that included *Nodularia spumigena* (*Cyanobacteria*), *Nostoc punctiforme* (*Cyanobacteria*), *Skermanella aerolata* (*Alphaproteobacteria*) and *Heterocystous Cyanobacterium* (*Cyanobacteria*) (Figure 3). Similarly, four indicator species were associated with farmland ecosystems (*Sinorhizobium* sp. TJ170 (*Alphaproteobacteria*), *Azospirillum zeae* (*Alphaproteobacteria*) and *Bradyrhizobium* sp. 61A124 (*Alphaproteobacteria*)), while *Bradyrhizobium* sp. ORS279 (*Alphaproteobacteria*), *Glucouncultured* sp. (*Alphaproteobacteria*) and *Azorhizobium doebereinerae* (*Alphaproteobacteria*) were found to be species-enriched indicators in the forest ecosystems. In the grassland ecosystems, *Bradyrhizobium* sp. ADD20940 (*Alphaproteobacteria*), *Derxia gummosa* (*Betaproteobacteria*), *Azospirillum halopraeferens* (*Alphaproteobacteria*), *Mesorhizobium septentrionale* (*Alphaproteobacteria*) and *Bradyrhizobium* sp. 58S1 (*Alphaproteobacteria*) were identified as indicator species. Also, four indicator species were found in mine ecosystems: *Nitrospirillum amazonense* (*Alphaproteobacteria*), *Bradyrhizobium* sp. ACT67986(*Alpha**proteobacteria*), *Bradyrhizobium* sp. TUXTLAS-18 (*Alphaproteobacteria*) and *Bradyrhizobium* sp. ACT67975(*Alphaproteobacteria*), and the four species-enriched taxa in riparian ecosystems included *Burkholderia xenovorans* (*Betaproteobacteria*), *Bradyrhizobium lablabi* (*Alphaproteobacteria*), *Methylocella tundrae* (*Alphaproteobacteria*) and *Leptothrix cholodnii* (*Betaproteobacteria*). In the tundra ecosystems, *Sphingomonas paucimobilis* (*Alphaproteobacteria*), *Bradyrhizobium* sp. IRBG228(*Alphaproteobacteria*), *Pseudomonas* sp. IPPW-3 (*Gamma**proteobacteria*) and *Mesorhizobium qingshengii* (*Alphaproteobacteria*) were the identified indicators. Together, these results indicate that *Bradyrhizobium* is a widely distributed diazotroph indicator genus followed by *Azospirillum* genus and *Cyanobacteria* phylum, even though investigated ecosystems included also unique indicator species.

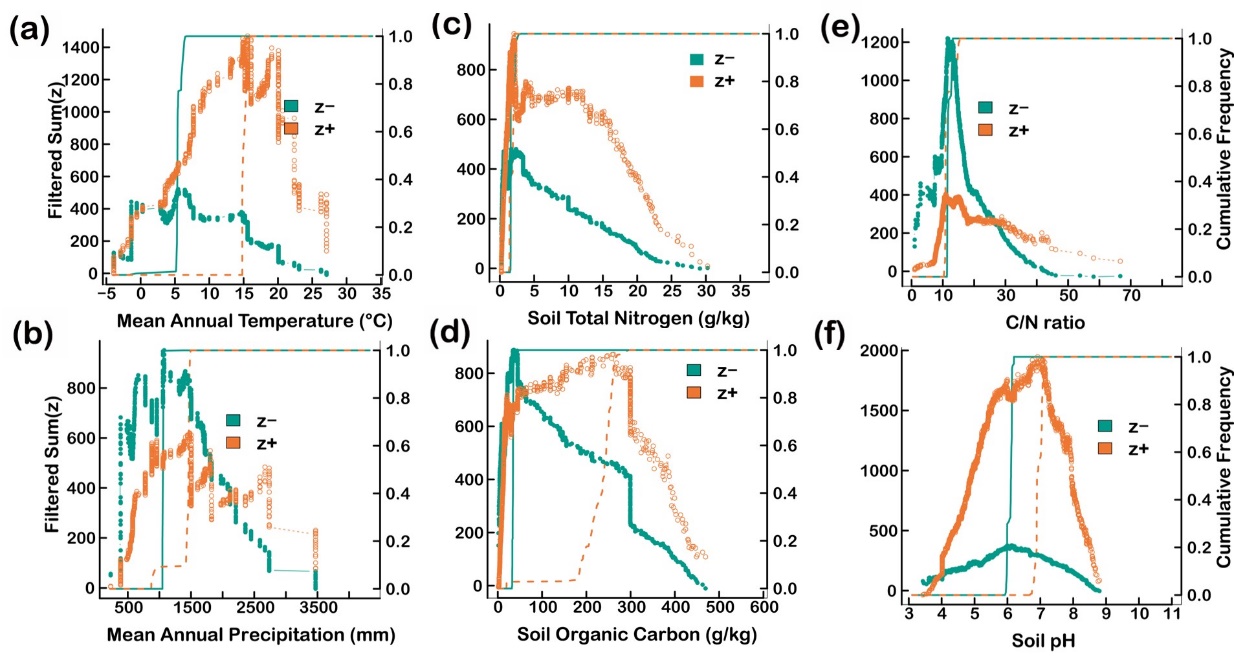


**Figure 3.** **Integrated indicator species analysis based on Indicator Value Analysis and random forest (RF) predictions.** Colored text on the left Y-axis represents the diazotroph Indicator Taxa identified by both methods in the corresponding ecosystems (legend for ecosystem colors is shown on top of the figure), while black text refers to indicator species predicted only by the RF model. The histogram sizes correspond to the relative abundance of each taxon in the corresponding ecosystem. The colored bar at the bottom of the figure shows the terrestrial ecosystem where the samples used in the analysis were collected from.

**Environmental change points and comprehensive diazotroph indicators species in terrestrial ecosystems**

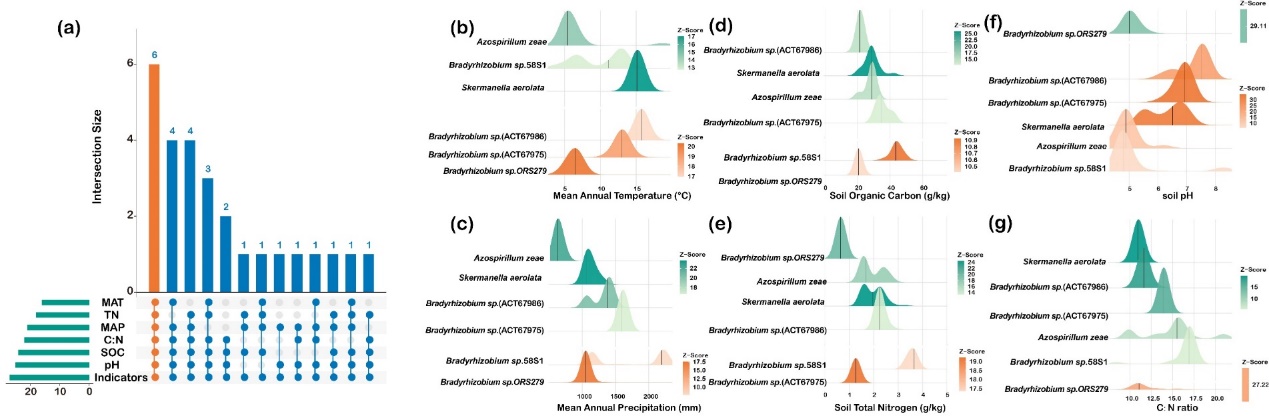
Threshold Indicator Taxa Analysis (TITAN) was used to detect changes in taxa distributions along environmental gradients. Specifically, we inferred environmental change points for the diazotroph community composition and taxon-specific contributions to community change along environmental gradients. This analysis enabled us to identify both positively and negatively responding taxa along key environmental parameters.

Since some of the environmental metadata was missing, environmental variables presented in at least half of the samples were selected for TITAN analysis. These variables included MAT (mean annual temperature), MAP (mean annual precipitation), soil pH, SOC (soil organic carbon), TN (total nitrogen of soil) and soil C:N ratio. The negative and positive threshold points for MAT occurred at 5.38 °C and 15.50 °C, respectively (Figure 4a), which means that diazotroph abundances declined below 5.30 °C and increased above 15.50 °C in our datasets. Moreover, 113 positively responding and 69 negatively responding diazotrophs were detected along with MAT (Appendix 4). With MAP, points of diazotroph decline and increase were 1066 mm and 1471 mm, respectively (Figure 4b) and 93 positively and 135 negatively responding taxa were identified. With TN, the positive and negative inflection points were similar at 2 g/kg (Figure 4c) with 131 positively and 100 negatively responding taxa. With SOC, the points of increase and decline were 34.6 g/kg and 261.4 g/kg, respectively (Figure 4d) and 96 positively and 144 negatively responding taxa were identified. In the case of C:N ratio, the points of increase and decline were 11.4 and 10.8, respectively (Figure 4e), with 56 taxa responding positively and 193 taxa negatively. Lastly, pH values of 6.1 and 6.9 represented the negative and the positive change threshold points, respectively (Figure 4f), with 188 positively and 61 negatively responding diazotroph taxa (Appendix 4). Together, this data suggests that diazotroph abundance and distribution were affected by all environmental properties that were included in the model.



**Figure 4.** **Threshold Indicator Taxa Analysis (TITAN) showing Sum(z-) and Sum(z+) values corresponding to points of decline and increase of diazotroph taxa.** The change points along the environmental parameters include mean annual temperature (a), mean annual precipitation (b), soil total nitrogen (c), soil organic carbon (d), C:N ratio (e) and soil pH (f). In all panels, left Y-axes show the filtered scores (Sum(z)) corresponding to all potential change points along the environmental gradient (filled circles), while right Y-axes show the cumulative frequency distribution of change points (solid line).

We next identified individual diazotroph species that strongly responded to above environmental gradients and could hence be considered as “environmental threshold indicator” species. A total of six such species were identified, which all were also predicted indicator species identified in all previous analyses (Figure 5a). These species included *Bradyrhizobium* sp. 58S1*, Bradyrhizobium* sp. ACT67986*, Bradyrhizobium* sp. ACT67975*, Bradyrhizobium* sp. ORS279*, Azospirillum zeae* and *Skermanella aerolata*, and they all showed different responses along with environmental gradients (Figures 5b-g). For example, the relative abundance of *Azospirillum zeae* increased with increasing pH (pH threshold point of 4.9), but decreased with increasing MAP, MAT SOC, TN and C:N ratio gradients (corresponding threshold points: 768 mm, 5.39 °C, 28.5 g/kg, 1.47 g/kg and 15.4, respectively). Even though the relative *Skermanella aerolata* abundance varied along different environmental gradient threshold points, it also increased along with increasing pH (pH threshold point of 6.9), and decreased with increasing MAP, MAT SOC, TN and C:N ratio gradients (corresponding threshold points: 1312 mm, 15.1 °C, 27.1 g/kg, 1.59 g/kg and 10.8, respectively). Interestingly, four identified *Bradyrhizobium* species responded differently along with environmental gradients. In case of MAP, *Bradyrhizobium* sp. ACT67986 *and Bradyrhizobium* sp. ACT67975abundances decreased, while the abundances of *Bradyrhizobium* sp. 58S1and *Bradyrhizobium* sp. ORS279increased with increasing precipitation (corresponding threshold points: 1066 mm, 1679 mm, 2276 mm and 1066 mm, respectively; Figure 5c). The four *Bradyrhizobium* strains responded differently not only to MAP, but also to the remaining five environmental factors. The relative abundance of *Bradyrhizobium* sp. 58S1increased with increasing SOC, TN and pH (corresponding threshold points: 44.53 g/kg, 3.65 g/kg and 4.59, respectively), but decreased with increasing MAT and C:N ratio gradients (corresponding threshold points: 6.75 °C, and 16.8, respectively). The relative abundance of *Bradyrhizobium* sp. ORS279increased with increasing MAT, SOC and C:N ratio (corresponding threshold points: 6.35 °C, 20.60 g/kg and 11.52, respectively), but decreased with increasing TN and pH gradients (corresponding threshold points: 0.70 g/kg, and 5.10, respectively). The relative abundance of *Bradyrhizobium* sp. ACT67975increased with increasing MAT, pH and TN (corresponding threshold points: 13.00 °C, 6.81 and 1.36 g/kg, respectively), but decreased with increasing SOC and C:N ratio gradients (corresponding threshold points: 33.58 g/kg, and 14.08, respectively). Finally, the relative abundance of *Bradyrhizobium* sp. ACT67986increased with increasing MAT and pH (corresponding threshold points: 15.50 °C and 7.68, respectively), but decreased with increasing SOC, C:N ratio and TN gradients (corresponding threshold points: 20.07 g/kg, 11.58 and 2.10 g/kg, respectively). Together, this analysis shows that diazotroph indicator species showed different responses along environmental gradients, suggesting that they are likely locally adapted to different terrestrial ecosystems.



**Figure 5. Integrated indicator species analysis linking diazotroph responses along six environmental gradients.** Panel (a) shows an upset plot representing the intersection of indicator species with different environmental characteristics based on TITAN analysis (every possible intersection represented by the bottom plot, and their occurrence shown on the top bar plot). Panels (b-g) show negative (green) and positive (orange) threshold values for six indicator diazotroph species along with mean annual temperature (b), mean annual precipitation (c), soil organic carbon (d), soil total nitrogen (e), C:N ratio (f) and soil pH (g). In (b-g), the peak points represent the species’ response threshold points.

**Discussion**

**Diazotroph community diversity patterns in typical terrestrial ecosystems**

Here, we used a custom-built *nifH* gene reference database and meta-analysis approach to study diazotroph diversity and community composition across terrestrial soil ecosystems. We also identified keystone diazotrophic species associated with specific ecosystems and determined how their abundances changed along with environmental characteristics. Large differences in the diazotroph community composition were found among the seven soil ecosystems included in this study (Figure 1d and 1e). This was not surprising as the roles and the relative importance of different diazotrophs are determined by their physiological adaptations and ecological properties, which closely match with their geographical distribution and environmental properties that vary between different terrestrial ecosystems (Landolfi et al. 2015; Wang et al. 2019; Zehr and Capone 2020).

Diazotrophs in mine ecosystems showed high diversity as indicated by the Chao1, Shannon and phylogenetic indices (Figure 1a, 1b and 1c). In general, mine ecosystems can be considered to represent harsh environments with elevated metal(loid) concentrations and nutrient deficiencies (Moynahan et al. 2002). It is commonly accepted that N fixation is more prevalent in N-deficient environments (Karthikeyan et al. 2019), which could partly explain high diazotroph diversity in mine ecosystems simply due to their ability to fix N. Moreover, Molybdenum (Mo) limitation to BNF has been reported in several terrestrial ecosystems (Gupta 1997; Hafner et al. 1992; Vieira et al. 1998a; Vieira et al. 1998b), and hence, relatively higher metal availability could have boosted BNF and increase diazotroph diversity in mines (Kabata-Pendias 2011; Sun et al. 2020). Moreover, it has been reported that diazotroph communities inhabiting mine tailings preferentially utilize inorganic electron donors, such as elemental sulfur (Sun et al. 2020). It is hence possible the sulfur oxidation pathway might be important for determining diazotroph success in mine ecosystems, highlighting the importance of linking other associated pathways with BNF.

High diazotroph diversity was also observed in forest ecosystems, which could be potentially explained by high carbon but low nitrogen availability. N fixation is energetically expensive and very sensitive to an adequate supply of energy (Ipata and Pesi 2015; Liang et al. 2016). In the forest ecosystems, tree litter provides the main source for organic carbon, playing an important role in maintaining forest ecosystems productivity, net carbon reserves as well as leading to the formation of soil organic matter, which can drive microbial community succession (Krishna and Mohan 2017; Zheng and Ding 1998). Canopy and litter N fixers are decoupled from soil N pool in forest (Menge & Hedin 2009), which creates N-poor conditions (plant tissues have high C:N ratios relative to decomposers) that favor N fixation despite soil N richness (Reed et al. 2008; Barron et al. 2009; Menge & Hedin 2009). Moreover, several previous studies have found leaf litter with a high labile C content but low total N content yielding resources with higher C:N that favored N fixers in the litter layer (Vitousek & Hobbie 2000; Perez et al. 2010). A bit surprisingly, high diazotroph diversity was also observed in farmland ecosystems where additional carbon (such as organic manure or crop straw) and nitrogen (fertilizers) are often incorporated into field soils by human activities (Fan et al. 2019; Guo et al. 2017; Kong et al. 2018). One explanation for this is that human activities and disturbance could increase the variation and total number of ecological niches in farmland soils, providing more ecological opportunities for diazotrophs (Hu et al. 2020) despite relatively high nitrogen availability. Moreover, it has been shown that diazotroph communities vary considerably between different farmlands and soil types (Han et al. 2019), and that organic fertilization could maintain diazotroph diversity regardless of N input (Feng et al. 2018). More studies considering the effect of microhabitat variability along with the nitrogen gradient are hence needed to better understand how nitrogen availability might promote diazotroph diversity instead of causing competitive exclusion of certain diazotrophic taxa.

The lowest diazotroph diversities were observed in crust and tundra ecosystems. Biological soil crust supply significant amounts of fixed nitrogen into terrestrial ecosystems worldwide (∼33 Tg y−1) (Hu et al. 2021). While it has been shown that crust diazotroph communities vary in their composition temporally, they often have relatively low diversity, consisting mainly of heterocystous *Cyanobacteria* (Yeager et al. 2012), which is in line with our findings. The highest TC content was identified in the tundra ecosystems (Figure S3). The plant-derived carbon has accumulated over hundreds to thousands of years in these ecosystems due to low temperatures, leading to saturated soils that have low levels of microbial decomposition of SOC (Lee et al. 2012; Pries et al. 2012). In general, low temperatures have a major negative effect on microbial activity (Stott et al. 1986), reducing prokaryote ability to sequester substrates from their environment (Nedwell 1999). Therefore, even though tundra ecosystems have high organic carbon availability, low temperature and microbial activity are likely to explain the observed low diazotroph diversity.

**Key abiotic characteristics associated with diazotrophic distribution**

We found that phylogenetically similar species responded to the same environmental factor differently and that the diazotroph’s responses to environmental variables varied considerably overall (Figure 3 and Figure 4). One explanation for this could be variation in soil nutrients (SOC, phosphorus, N availability, and C:N ratio) and soil pH (Hsu and Buckley 2009; Wang et al. 2017; Wang et al. 2016). For example, the positive and negative thresholds points for organic carbon were 34.6 g/kg and 261.4 g/kg, respectively (Figure 3d) and one explanation for such a wide range of variation could be that diazotrophs have adapted to carbon-rich and carbon-poor environments, respectively. In line with this reasoning, diazotrophs affiliated with *Alphaproteobacteria* and *Betaproteobacteria* are usually classified as copiotrophs (Trivedi et al. 2013) and might benefit from high resource availability that can support their growth and functioning (Fierer et al. 2007). In contrast, diazotrophs belonging to *Deltaproteobacteria* are often considered as oligotrophs (Trivedi et al. 2013) and might be better adapted to BNF in low-nutrient environments (Koch 2001). In the case of soil TN threshold, a value close to 2.0 g/kg was identified as the most dramatic threshold point for diazotrophs across all terrestrial ecosystems (Figure 4c, Appendix 4). This suggests that diazotroph diversity is very sensitive to TN, where low and high TN contents might reduce the diversity of diazotrophs despite the type of terrestrial ecosystem. In case of C:N ratio, a value close to 11 was identified as an important threshold point (Figure 4e). To acquire carbon and nitrogen, soil microorganisms require a diet with a C:N ratio near to 24:1, with 16 parts of carbon used for energy acquisition and 8 parts for the cell maintenance. The lower the C:N ratio, more rapidly N will be released into the soil for immediate crop use (Watson et al. 2002). In contrast, if the C:N ratio exceeds 35, N immobilization will be observed, while a ratio between 20 to 30 will result in an equilibrium state between mineralization and immobilization (Brust 2019). The C:N ratio of natural soils are most often close to values of 10. This suggests that diazotrophs might have adapted their C:N ratios use according to the values typically observed in most terrestrial ecosystems. However, more research is required to understand why a threshold value closer to 11 is so important for diazotroph diversity.

In addition to nutrients, soil pH, temperature and precipitation play key roles in influencing the diversity and community composition of prokaryotes including diazotrophs (Averill et al. 2016; Evans and Wallenstein 2012; Razanamalala et al. 2018; Xue et al. 2016; Zhou et al. 2016). The pH affects microbial growth and survival and is a key player in redox reactions, activity of extracellular enzymes and microbial metabolism (Jin and Kirk 2018). It is known that neutral environments are more suitable for microbial growth, metabolism and attachment to negatively charged particles, while low pH may affect the activity of negatively charged biofilms (Aoyagi et al. 2017). As a result, microbial richness has been shown to increase up to a near-neutral pH at the global scale (Thompson et al. 2017). In line with this, we found that diazotroph diversity and relative abundances peaked at neutral pH near 7. Diazotroph diversity could also be shaped by temperature and precipitation (Fuhrman et al. 2008). For example, diazotrophs have been shown to respond positively at 15.50 °C (Thompson et al. 2017), while bacterial biomass has been shown to increase with precipitation across a range of ecosystems (Griffiths et al. 2011). We found that indicator diazotroph species were associated with a wide range of both temperature and precipitation variation, which reflects diazotrophs’ ability to adapt to their local environmental weather conditions. Furthermore, diazotroph distribution could be driven by interactions between multiple environmental variables, making it challenging to disentangle the relative importance of different factors, especially when using a meta-analysis approach that combines data across several studies. For example, temperature, precipitation, and N deposition are relatively high in the subtropics, but the soils tend to be N-rich and P-deficient which leads to rapid mineralization. In contrast, temperate soils are generally rich in organic C and phosphorus but relatively deficient in N (Xu et al. 2017; Zhang et al. 2018). As such, temperature and precipitation may affect the diversity of N-fixing bacteria together through co-correlations with soil factors, and future research should focus on exploring the multifactorial environmental effects on diazotroph communities. Together, our findings suggest that while diazotrophs have likely adapted to a range of environmental niches, it is important to focus on specific biomes in the future to understand in what extent the relative abundance and diversity of *nifH*-associated taxa are explained by certain specific environmental parameters at the local level.

**Typical diazotrophic indicator species across terrestrial ecosystems**

Members of the *Bradyrhizobium* are known facultative N fixers (Rivas et al. 2009) and has been shown to play dominant roles across many plant types (Mirza et al. 2014). They are also known to occupy wide range niches and are can survive across diverse environments (Bottomley et al. 1991). It was hence not a surprise that diazotrophic *Bradyrhizobium* species were most often predicted to be diazotroph indicator species regardless of the terrestrial ecosystem type. For instance, *Bradyrhizobium* sp. ORS279, *Bradyrhizobium* sp. 61A124 and *Bradyrhizobium* sp. IRBG228 were found to be indicator species in forest, farmland and tundra ecosystems, respectively (Figure 3). In contrast, the genus *Azospirillum* often forms biofilms (Herath et al. 2015; Rinaudi and Giordano 2010), which could allow it to persist in a range of different environmental conditions across terrestrial ecosystems. Yet, the majority of diazotrophs belonged to the “above prediction” and “below prediction” groups (Figure S2), which suggests that their presence could be driven by variation in specific environmental conditions. For example, diazotrophs in “above prediction” groups could be positively, while diazotrophs in “below prediction” groups could be negatively affected by specific environmental parameters. The all six comprehensive predictor species belonged to either the “above” or the “below” prediction groups (Appendix 3), which indicates that they were all strongly but somewhat differently influenced by the environmental factors. Together, these analyses suggest that it might be possible to identify diazotroph indicator species associated to certain terrestrial ecosystems, and to identify key environmental conditions that are positively associated with non-symbiotic N-fixation.

**Conclusions and caveats**

Collectively, this study shows that diazotroph community diversity, structure, and distribution varies considerably across terrestrial ecosystems on a continental scale. *Azospirillum zeae*, *Skermanella aerolata* and four *Bradyrhizobium species* were identified as comprehensive indicators for distinguishing between different ecosystem types and environmental characteristics that favour N fixation. While this suggests that diazotrophs are widely adapted to a variety of environments, more work is needed to identify between generalist and specialist taxa and their associations with different environmental niches. Moreover, a better understanding on interactions and collinearity between environmental factors on diazotroph distribution and activity is required, which is hard to control in our meta-analysis, which pools together data from multiple different studies. Moreover, the patterns observed in this study do not necessary suggest a causality between diazotroph distribution and environmental factors, and hence, experimental manipulations and finer scale comparative analyses within individual terrestrial ecosystem studies are required. For example, it would be interesting to explore in the future if certain farm systems are more often associated with diazotrophs and BNF and how their abundance is affected by other environmental variables such as N input. Also, our sequence-based analysis cannot unequivocally conclude if *nifH* sequence presence is indicative of active BNF or the presence of inactive or dead cells. While it has been reported that *Bradyrhizobium* spp. are capable of BNF even at non-symbiotic state (Terakado-Tonooka and Ohwaki 2013; Wongdee et al. 2018), functional based analyses are required to link *nifH* gene presence with active BNF using metatranscriptomics or RT-qPCR. Moreover, validation of *nifH* sequence-based taxonomic predictions with 16S rRNA data would be useful to better understand the link between phylogenetic and functional diversity. In summary, our results provide insights into the overall distribution and underlying environmental determinants of diazotroph communities, providing a basis for improving BNF in terrestrial ecosystems globally and locally.

**Acknowledgements**

We are grateful for many graduate students and staffs who involved in collecting data and providing helpful suggestions but are not listed as co-authors. This work was supported by the National Natural Science Foundation of China (41977080, 31902114, U2003210), the Natural Science Foundation of Jiangsu Province (BK20190543), the China Postdoctoral Science Foundation (2019M651861), the Young Elite Scientists Sponsorship Program by CAST (2019QNRC001), the Innovative Research Team Development Plan of the Ministry of Education of China (Grant No. IRT\_17R56), and the Fundamental Research Funds for the Central Universities (Grant No. KYT201802).

**Authors' contributions**

N.L., S.W.G. and Q.R.S. conceived and designed the study. C.Z., L.L. and Q.C.X. performed the analyses and drafted the manuscript. V-P.F., S.W.G. and Q.R.S. made plentiful valuable comments for the manuscript. All the authors reviewed and approved the manuscript.

**Competing interests**

The authors declare that they have no conflict of interest.

**Data availability statement**

All occurrence data are available from the published articles or the National Centre for Biotechnology Information (NCBI) SRA database, and the DOIs for all referenced literature or the data links are provided in the Supporting Information.

**References:**

Affourtit, J., Zehr, J., and Paerl, H. (2001) Distribution of nitrogen-fixing microorganisms along the Neuse River Estuary, North Carolina. Microb Ecol 41: 114-123.

Alazard, D. (1990) Nitrogen fixation in pure culture by rhizobia isolated from stem nodules of tropical Aeschynomene species. FEMS Microbiol Lett 68: 177-182.

Averill, C., Waring, B.G., and Hawkes, C.V. (2016) Historical precipitation predictably alters the shape and magnitude of microbial functional response to soil moisture. Global Change Biol 22: 1957-1964.

Baker, M.E., and King, R.S. (2010) A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution 1: 25-37.

Baker, M.E., King, R.S., Kahle, D., and Kahle, M.D. (2015) Package ‘TITAN2’. In.

Barron, A.R., Wurzburger, N., Bellenger, J.P., Wright, S.J., Kraepiel, A.M., and Hedin, L.O. (2009) Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. Nature Geoscience 2: 42-45.

Bottomley, P.J., Maggard, S.P., Leung, K., and Busse, M.D. (1991) Importance of saprophytic competence for introduced rhizobia. In The Rhizosphere and Plant Growth: Springer, pp. 135-140.

Brust, G.E. (2019) Management strategies for organic vegetable fertility. In Safety and practice for organic food: Elsevier, pp. 193-212.

Chalk, P.M., He, J.-Z., Peoples, M.B., and Chen, D. (2017) 15N2 as a tracer of biological N2 fixation: a 75-year retrospective. Soil Biol Biochem 106: 36-50.

Che, R., Deng, Y., Wang, F., Wang, W., Xu, Z., Hao, Y. et al. (2018) Autotrophic and symbiotic diazotrophs dominate nitrogen-fixing communities in Tibetan grassland soils. Sci Total Environ 639: 997-1006.

Cheng, Q. (2008) Perspectives in biological nitrogen fixation research. Journal of integrative plant biology 50: 786-798.

Cleveland, C.C., Townsend, A.R., Schimel, D.S., Fisher, H., Howarth, R.W., Hedin, L.O. et al. (1999) Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. Global Change Biol 13: 623-645.

Dreyfus, B., Elmerich, C., and Dommergues, Y. (1983) Free-living Rhizobium strain able to grow on N2 as the sole nitrogen source. Appl Environ Microbiol 45: 711-713.

Dreyfus, B., Garcia, J.-L., and Gillis, M. (1988) Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from Sesbania rostrata. Int J Syst Evol Microbiol 38: 89-98.

Dufrêne, M., and Legendre, P. (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr 67: 345-366.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10: 996-998.

Evans, S.E., and Wallenstein, M.D. (2012) Soil microbial community response to drying and rewetting stress: does historical precipitation regime matter? Biogeochemistry 109: 101-116.

Fan, K., Delgado-Baquerizo, M., Guo, X., Wang, D., Wu, Y., Zhu, M. et al. (2019) Suppressed N fixation and diazotrophs after four decades of fertilization. Microbiome 7: 1-10.

Feng, M., Adams, J.M., Fan, K., Shi, Y., Sun, R., Wang, D. et al. (2018) Long-term fertilization influences community assembly processes of soil diazotrophs. Soil Biol Biochem 126: 151-158.

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

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. Proceedings of the National Academy of Sciences 105: 7774-7778.

Gaby, J.C., and Buckley, D.H. (2011) A global census of nitrogenase diversity. Environ Microbiol 13: 1790-1799.

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P. et al. (2004) Nitrogen cycles: past, present, and future. Biogeochemistry 70: 153-226.

Griffiths, R.I., Thomson, B.C., James, P., Bell, T., Bailey, M., and Whiteley, A.S. (2011) The bacterial biogeography of British soils. Environ Microbiol 13: 1642-1654.

Guo, J., Ling, N., Chen, H., Zhu, C., Kong, Y., Wang, M. et al. (2017) Distinct drivers of activity, abundance, diversity and composition of ammonia-oxidizers: evidence from a long-term field experiment. Soil Biol Biochem 115: 403-414.

Gupta, U.C. (1997) Molybdenum in agriculture: Cambridge University Press.

Gupta, V., Kroker, S., Hicks, M., Davoren, C., Descheemaeker, K., and Llewellyn, R. (2014) Nitrogen cycling in summer active perennial grass systems in South Australia: non-symbiotic nitrogen fixation. Crop and Pasture Science 65: 1044-1056.

Hafner, H., Ndunguru, B., Bationo, A., and Marschner, H. (1992) Effect of nitrogen, phosphorus and molybdenum application on growth and symbiotic N2-fixation of groundnut in an acid sandy soil in Niger. Fertilizer Research 31: 69-77.

Hall, J.R., Mitchell, K.R., Jackson-Weaver, O., Kooser, A.S., Cron, B.R., Crossey, L.J., and Takacs-Vesbach, C.D. (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol 74: 4910-4922.

Hamelin, J., Fromin, N., Tarnawski, S., Teyssier‐Cuvelle, S., and Aragno, M. (2002) nifH gene diversity in the bacterial community associated with the rhizosphere of Molinia coerulea, an oligonitrophilic perennial grass. Environ Microbiol 4: 477-481.

Hamilton, T.L., Boyd, E.S., and Peters, J.W. (2011) Environmental constraints underpin the distribution and phylogenetic diversity of nifH in the Yellowstone geothermal complex. Microb Ecol 61: 860-870.

Han, L.-L., Wang, Q., Shen, J.-P., Di, H.J., Wang, J.-T., Wei, W.-X. et al. (2019) Multiple factors drive the abundance and diversity of the diazotrophic community in typical farmland soils of China. FEMS Microbiol Ecol 95: fiz113.

Hayden, H.L., Drake, J., Imhof, M., Oxley, A.P., Norng, S., and Mele, P.M. (2010) The abundance of nitrogen cycle genes amoA and nifH depends on land-uses and soil types in South-Eastern Australia. Soil Biol Biochem 42: 1774-1783.

Herath, I., Kumarathilaka, P., Navaratne, A., Rajakaruna, N., and Vithanage, M. (2015) Immobilization and phytotoxicity reduction of heavy metals in serpentine soil using biochar. J Soils Sed 15: 126-138.

Herridge, D.F., Peoples, M.B., and Boddey, R.M. (2008) Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil 311: 1-18.

Hollander, M., Wolfe, D.A., and Chicken, E. (2013) Nonparametric statistical methods: John Wiley & Sons.

Hsu, S.-F., and Buckley, D.H. (2009) Evidence for the functional significance of diazotroph community structure in soil. The ISME journal 3: 124-136.

Hu, J., Richwine, J.D., Keyser, P.D., Li, L., Yao, F., Jagadamma, S., and DeBruyn, J.M. (2021) Nitrogen Fertilization and Native C4 Grass Species Alter Abundance, Activity, and Diversity of Soil Diazotrophic Communities. Frontiers in microbiology 12.

Hu, X., Liang, A., Yao, Q., Liu, Z., Yu, Z., Wang, G., and Liu, J. (2020) Ridge tillage improves soil properties, sustains diazotrophic communities, and enhances extensively cooperative interactions among diazotrophs in a clay loam soil. Frontiers in microbiology 11: 1333.

Ipata, P.L., and Pesi, R. (2015) What is the true nitrogenase reaction? A guided approach. Biochemistry and Molecular Biology Education 43: 142-144.

Kabata-Pendias, A. (2011) Trace elements in soils and plants.,(4th edn)(CRC Press: Boca Raton, FL, USA).

Karthikeyan, S., Rodriguez-R, L.M., Heritier-Robbins, P., Kim, M., Overholt, W.A., Gaby, J.C. et al. (2019) “Candidatus Macondimonas diazotrophica”, a novel gammaproteobacterial genus dominating crude-oil-contaminated coastal sediments. The ISME journal 13: 2129-2134.

King, R.S., and Baker, M.E. (2010) Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. J N Am Benthol Soc 29: 998-1008.

Koch, A.L. (2001) Oligotrophs versus copiotrophs. Bioessays 23: 657-661.

Kong, Y., Zhu, C., Ruan, Y., Luo, G., Wang, M., Ling, N. et al. (2018) Are the microbial communities involved in glucose assimilation in paddy soils treated with different fertilization regimes for three years similar? J Soils Sed 18: 2476-2490.

Krishna, M., and Mohan, M. (2017) Litter decomposition in forest ecosystems: a review. Energy, Ecology and Environment 2: 236-249.

Landolfi, A., Koeve, W., Dietze, H., Kähler, P., and Oschlies, A. (2015) A new perspective on environmental controls of marine nitrogen fixation. Geophys Res Lett 42: 4482-4489.

Langlois, R.J., LaRoche, J., and Raab, P.A. (2005) Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. Appl Environ Microbiol 71: 7910-7919.

LeBauer, D.S., and Treseder, K.K. (2008) Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. Ecology 89: 371-379.

Lee, H., Schuur, E.A., Inglett, K.S., Lavoie, M., and Chanton, J.P. (2012) The rate of permafrost carbon release under aerobic and anaerobic conditions and its potential effects on climate. Global Change Biol 18: 515-527.

Lee, K.-B., De Backer, P., Aono, T., Liu, C.-T., Suzuki, S., Suzuki, T. et al. (2008) The genome of the versatile nitrogen fixer Azorhizobium caulinodans ORS571. BMC Genomics 9: 1-14.

Lee, T.K., Han, I., Kim, M.S., Seong, H.J., Kim, J.-S., and Sul, W.J. (2019) Characterization of a nifH-harboring Bacterial Community in the Soil-limited Gotjawal Forest. Front Microbiol 10: 1858.

Levy-Booth, D.J., Prescott, C.E., and Grayston, S.J. (2014) Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. Soil Biol Biochem 75: 11-25.

Liang, Y., Pan, F., He, X., Chen, X., and Su, Y. (2016) Effect of vegetation types on soil arbuscular mycorrhizal fungi and nitrogen-fixing bacterial communities in a karst region. Environ Sci Pollut R 23: 18482-18491.

Liaw, A., and Wiener, M. (2002) Classification and regression by randomForest. R news 2: 18-22.

Matson, A.L., Corre, M.D., Burneo, J.I., and Veldkamp, E. (2015) Free-living nitrogen fixation responds to elevated nutrient inputs in tropical montane forest floor and canopy soils of southern Ecuador. Biogeochemistry 122: 281-294.

Menge, D.N., and Hedin, L.O. (2009) Nitrogen fixation in different biogeochemical niches along a 120 000‐year chronosequence in New Zealand. Ecology 90: 2190-2201.

Mirza, B.S., Potisap, C., Nüsslein, K., Bohannan, B.J., and Rodrigues, J.L. (2014) Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. Appl Environ Microbiol 80: 281-288.

Monk, W.A., Compson, Z.G., Armanini, D.G., Orlofske, J.M., Curry, C.J., Peters, D.L. et al. (2018) Flow velocity–ecology thresholds in Canadian rivers: A comparison of trait and taxonomy‐based approaches. Freshwat Biol 63: 891-905.

Moynahan, O.S., Zabinski, C.A., and Gannon, J.E. (2002) Microbial community structure and carbon‐utilization diversity in a mine tailings revegetation study. Restor Ecol 10: 77-87.

Nedwell, D.B. (1999) Effect of low temperature on microbial growth: lowered affinity for substrates limits growth at low temperature. FEMS Microbiol Ecol 30: 101-111.

Nguyen, T.H.T., Forio, M.A.E., Boets, P., Lock, K., Damanik Ambarita, M.N., Suhareva, N. et al. (2018) Threshold responses of macroinvertebrate communities to stream velocity in relation to hydropower dam: A case study from the Guayas river basin (Ecuador). Water 10: 1195.

Ning, D., Deng, Y., Tiedje, J.M., and Zhou, J. (2019) A general framework for quantitatively assessing ecological stochasticity. Proceedings of the National Academy of Sciences 116: 16892-16898.

Peoples, M.B., and Craswell, E.T. (1992) Biological nitrogen fixation: investments, expectations and actual contributions to agriculture. Plant Soil 141: 13-39.

Pérez, C.A., Carmona, M.R., and Armesto, J.J. (2010) Non‐symbiotic nitrogen fixation during leaf litter decomposition in an old‐growth temperate rain forest of Chiloé Island, southern Chile: Effects of single versus mixed species litter. Austral Ecol 35: 148-156.

Pries, C.E.H., Schuur, E.A., and Crummer, K.G. (2012) Holocene carbon stocks and carbon accumulation rates altered in soils undergoing permafrost thaw. Ecosystems 15: 162-173.

Razanamalala, K., Razafimbelo, T., Maron, P.-A., Ranjard, L., Chemidlin, N., Lelièvre, M. et al. (2018) Soil microbial diversity drives the priming effect along climate gradients: a case study in Madagascar. The ISME journal 12: 451-462.

Reed, S.C., Cleveland, C.C., and Townsend, A.R. (2008) Tree species control rates of free‐living nitrogen fixation in a tropical rain forest. Ecology 89: 2924-2934.

Reed, S.C., Cleveland, C.C., and Townsend, A.R. (2011) Functional ecology of free-living nitrogen fixation: a contemporary perspective. Annual review of ecology, evolution, and systematics 42: 489-512.

Reed, S.C., Cleveland, C.C., and Townsend, A.R. (2013) Relationships among phosphorus, molybdenum and free-living nitrogen fixation in tropical rain forests: results from observational and experimental analyses. Biogeochemistry 114: 135-147.

Rinaudi, L.V., and Giordano, W. (2010) An integrated view of biofilm formation in rhizobia. FEMS Microbiol Lett 304: 1-11.

Rivas, R., Martens, M., De Lajudie, P., and Willems, A. (2009) Multilocus sequence analysis of the genus Bradyrhizobium. Syst Appl Microbiol 32: 101-110.

Rosen, C.J., and Allan, D.L. (2007) Exploring the benefits of organic nutrient sources for crop production and soil quality. HortTechnology 17: 422-430.

Son, Y. (2001) Non-symbiotic nitrogen fixation in forest ecosystems. Ecol Res 16: 183-196.

Stott, D., Elliott, L., Papendick, R., and Campbell, G. (1986) Low temperature or low water potential effects on the microbial decomposition of wheat residue. Soil Biol Biochem 18: 577-582.

Sun, X., Kong, T., Häggblom, M.M., Kolton, M., Li, F., Dong, Y. et al. (2020) Chemolithoautotropic diazotrophy dominates the nitrogen fixation process in mine tailings. Environ Sci Technol 54: 6082-6093.

Tarnocai, C., Canadell, J., Schuur, E.A., Kuhry, P., Mazhitova, G., and Zimov, S. (2009) Soil organic carbon pools in the northern circumpolar permafrost region. Global biogeochemical cycles 23.

Team, R.C. (2013) R: A language and environment for statistical computing.

Terakado-Tonooka, J., Fujihara, S., and Ohwaki, Y. (2013) Possible contribution of *Bradyrhizobium* on nitrogen fixation in sweet potatoes. Plant Soil 367: 639-650.

Trivedi, P., Anderson, I.C., and Singh, B.K. (2013) Microbial modulators of soil carbon storage: integrating genomic and metabolic knowledge for global prediction. Trends Microbiol 21: 641-651.

Tu, Q., Zhou, X., He, Z., Xue, K., Wu, L., Reich, P. et al. (2016) The diversity and co-occurrence patterns of N 2-fixing communities in a CO2-enriched grassland ecosystem. Microb Ecol 71: 604-615.

Turk, K.A., Rees, A.P., Zehr, J.P., Pereira, N., Swift, P., Shelley, R. et al. (2011) Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. The ISME journal 5: 1201-1212.

Vieira, R., Cardoso, E., Vieira, C., and Cassini, S. (1998a) Foliar application of molybdenum in common beans. I. Nitrogenase and reductase activities in a soil of high fertility. J Plant Nutr 21: 169-180.

Vieira, R., Vieira, C., Cardoso, E., and Mosquim, P. (1998b) Foliar application of molybdenum in common bean. II. Nitrogenase and nitrate reductase activities in a soil of low fertility. J Plant Nutr 21: 2141-2151.

Vitousek, P.M., and Howarth, R.W. (1991) Nitrogen limitation on land and in the sea: how can it occur? Biogeochemistry 13: 87-115.

Vitousek, P.M., and Hobbie, S. (2000) Heterotrophic nitrogen fixation in decomposing litter: patterns and regulation. Ecology 81: 2366-2376.

Vitousek, P.M., Menge, D.N., Reed, S.C., and Cleveland, C.C. (2013) Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B: Biological Sciences 368: 20130119.

Wakelin, S., Gupta, V., and Forrester, S. (2010) Regional and local factors affecting diversity, abundance and activity of free-living, N2-fixing bacteria in Australian agricultural soils. Pedobiologia 53: 391-399.

Wang, C., Zheng, M., Song, W., Wen, S., Wang, B., Zhu, C., and Shen, R. (2017) Impact of 25 years of inorganic fertilization on diazotrophic abundance and community structure in an acidic soil in southern China. Soil Biol Biochem 113: 240-249.

Wang, J., Zhang, D., Zhang, L., Li, J., Raza, W., Huang, Q., and Shen, Q. (2016) Temporal variation of diazotrophic community abundance and structure in surface and subsoil under four fertilization regimes during a wheat growing season. Agric, Ecosyst Environ 216: 116-124.

Wang, Q., Quensen, J.F., Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., and Cole, J.R. (2013) Ecological patterns of nifH genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. MBio 4.

Wang, W.-L., Moore, J.K., Martiny, A.C., and Primeau, F.W. (2019) Convergent estimates of marine nitrogen fixation. Nature 566: 205-211.

Watson, C., Atkinson, D., Gosling, P., Jackson, L., and Rayns, F. (2002) Managing soil fertility in organic farming systems. Soil use and management 18: 239-247.

Wongdee, J., Boonkerd, N., Teaumroong, N., Tittabutr, P., and Giraud, E. (2018) Regulation of nitrogen fixation in *Bradyrhizobium* sp. strain DOA9 involves two distinct *NifA* regulatory proteins that are functionally redundant during symbiosis but not during free-living growth. Frontiers in microbiology 9: 1644.

Wickham, H. (2016) ggplot2-Elegant Graphics for Data Analysis. Springer International Publishing. Cham, Switzerland.

Xiao, D., Xiao, L., Che, R., Tan, Y., Liu, X., Yang, R. et al. (2020) Phosphorus but not nitrogen addition significantly changes diazotroph diversity and community composition in typical karst grassland soil. Agric, Ecosyst Environ 301: 106987.

Xu, S., Wang, Y., Huang, B., Wei, Z.-B., Miao, A.-J., and Yang, L.-Y. (2015) Nitrogen and phosphorus limitation of phytoplankton growth in different areas of Lake Taihu, China. J Freshwat Ecol 30: 113-127.

Xu, Z., Yu, G., Zhang, X., He, N., Wang, Q., Wang, S. et al. (2017) Soil enzyme activity and stoichiometry in forest ecosystems along the North-South Transect in eastern China (NSTEC). Soil Biol Biochem 104: 152-163.

Xue, K., Yuan, M.M., Shi, Z.J., Qin, Y., Deng, Y., Cheng, L. et al. (2016) Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. Nature Climate Change 6: 595-600.

Yeager, C.M., Kuske, C.R., Carney, T.D., Johnson, S.L., Ticknor, L.O., and Belnap, J. (2012) Response of biological soil crust diazotrophs to season, altered summer precipitation, and year-round increased temperature in an arid grassland of the Colorado Plateau, USA. Frontiers in Microbiology 3: 358.

Zehr, J.P., and Capone, D.G. (2020) Changing perspectives in marine nitrogen fixation. Science 368.

Zhang, X., Yang, Y., Zhang, C., Niu, S., Yang, H., Yu, G. et al. (2018) Contrasting responses of phosphatase kinetic parameters to nitrogen and phosphorus additions in forest soils. Funct Ecol 32: 106-116.

Zheng, Y., and Ding, Y. (1998) Effect of mixed forests of Chinese fir and tsoong’s tree on soil properties. Pedosphere 8: 161-168.

Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D. et al. (2016) Temperature mediates continental-scale diversity of microbes in forest soils. Nature communications 7: 1-10.

**Figure list**

**Figure 1. Differences in diazotroph community diversity and composition across soil ecosystems.** Panels (a-c) show the α-diversity indexes of diazotroph communities from seven terrestrial ecosystems (including Chao1 index (a), Shannon diversity (b) and phylogenetic diversity (c). Panel (d) shows phylum-level composition of diazotroph communities. Panel (e) shows Principal Coordinates Analysis (PCoA) of diazotroph community composition based on species matrices from seven terrestrial ecosystems. One-way PERMANOVA was used to analyses the effects of ecosystem-types on community structure of diazotroph communities.

**Figure 2. Diazotroph indicator species analysis based on Random Forest (RF) model across different soil ecosystems**. A total of 32 bacterial species were identified by applying random-forest classification of the relative abundance of the diazotrophs. Panels (a-b) show predicted indicator species that are ranked in descending order of importance following the accuracy of the model (70% of data used for RF model). (c) The verification of RF model results (30% of remaining data) showing correctly (shown on given ecosystem colors) and incorrectly (black color) predicted indicator diazotrophs in seven ecosystems. In (c), each square corresponds to an individual sampling location.

**Figure 3.** **Integrated indicator species analysis based on Indicator Value Analysis and random forest (RF) predictions.** Colored text on the left Y-axis represents the diazotroph Indicator Taxa identified by both methods in the corresponding ecosystems (legend for ecosystem colors is shown on top of the figure), while black text refers to indicator species predicted only by the RF model. The histogram sizes correspond to the relative abundance of each taxon in the corresponding ecosystem. The colored bar at the bottom of the figure shows the terrestrial ecosystem where the samples used in the analysis were collected from.

**Figure 4.** **Threshold Indicator Taxa Analysis (TITAN) showing Sum(z-) and Sum(z+) values corresponding to points of decline and increase of diazotroph taxa.** The change points along the environmental parameters include mean annual temperature (a), mean annual precipitation (b), soil total nitrogen (c), soil organic carbon (d), C:N ratio (e) and soil pH (f). In all panels, left Y-axes show the filtered scores (Sum(z)) corresponding to all potential change points along the environmental gradient (filled circles), while right Y-axes show the cumulative frequency distribution of change points (solid line).

**Figure 5.** Integrated indicator species analysis linking diazotroph responses along six environmental gradients. Panel (a) shows an upset plot representing the intersection of indicator species with different environmental characteristics based on TITAN analysis (every possible intersection represented by the bottom plot, and their occurrence shown on the top bar plot). Panels (b-g) show negative (green) and positive (orange) threshold values for six indicator diazotroph species along with mean annual temperature (b), mean annual precipitation (c), soil organic carbon (d), soil total nitrogen (e), C:N ratio (f) and soil pH (g). In (b-g), the peak points represent the species’ response threshold points.

**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Figure S1** The global distribution of experiments providing *nifH* gene sequence data that were included in the meta-analysis. Different colors indicate seven types of terrestrial ecosystems and the pie chart represents the percentage of studies included in each ecosystem type.

**Figure S2** Fit of the neutral models for diazotrophic community data with a diazotrophic metacommunity as the source. Panel (a) shows the species that occurred more frequently than predicted by the model in green, while those species that occurred less frequently than predicted are shown in orange. Blue dashed lines represent 95% confidence intervals around the model predictions and the species that fell within the confidence intervals were considered as neutrally distributed. R2 values present the goodness of fit of the neutral model, ranging from 0 (no fit) to 1 (perfect fit). Panels (b-c) show the cumulative relative abundance and taxonomic distribution of species, respectively, in three diazotrophic ecosystems classified in three categories: above prediction, below prediction and neutrally distributed.

**Figure S3** The mean environmental data for all terrestrial ecosystems based on 39 independent studies used in the meta-analysis. The boxplots in panels (a-f) show mean annual temperature (MAT), mean annual precipitation (MAP), pH, moisture, soil organic carbon (SOC), total nitrogen (TN), carbon nitrogen ratio (C:N), total phosphorus (TP), total potassium (TK), ammonium nitrogen (NH4+-N) and nitrate nitrogen (NO3--N), respectively.

**Appendix 1** Information of collected samples.

**Appendix 2** Annotation information for nitrogen-fixing prokaryotes.

**Appendix 3** The results of neutral models for Farmland, Forest and Grassland ecosystems.

**Appendix 4** The results of Titan analysis.

**Appendix 5** The results of Indicator analysis for environmental factors.