RESEARCH

Open Access

Global soil metagenomics reveals distribution and predominance of *Deltaproteobacteria* in nitrogen-fixing microbiome

Yoko Masuda[1,2*†], Kazumori Mise[3*†], Zhenxing Xu[1], Zhengcheng Zhang[1], Yutaka Shiratori[4], Keishi Senoo[1,2] and Hideomi Itoh[3*]

Abstract

Background Biological nitrogen fixation is a fundamental process sustaining all life on earth. While distribution and diversity of N_2-fixing soil microbes have been investigated by numerous PCR amplicon sequencing of nitrogenase genes, their comprehensive understanding has been hindered by lack of *de facto* standard protocols for amplicon surveys and possible PCR biases. Here, by fully leveraging the planetary collections of soil shotgun metagenomes along with recently expanded culture collections, we evaluated the global distribution and diversity of terrestrial diazotrophic microbiome.

Results After the extensive analysis of 1,451 soil metagenomic samples, we revealed that the *Anaeromyxobacteraceae* and *Geobacteraceae* within *Deltaproteobacteria* are ubiquitous groups of diazotrophic microbiome in the soils with different geographic origins and land usage types, with particular predominance in anaerobic soils (paddy soils and sediments).

Conclusion Our results indicate that *Deltaproteobacteria* is a core bacterial taxon in the potential soil nitrogen fixation population, especially in anaerobic environments, which encourages a careful consideration on deltaproteobacterial diazotrophs in understanding terrestrial nitrogen cycling.

Keywords microbial community, soil microbiome, metagenomics, nitrogen fixation

†Yoko Masuda and Kazumori Mise contributed equally to this work.

*Correspondence:
Yoko Masuda
yokomasuda@g.ecc.u-tokyo.ac.jp; ygigico@gmail.com
Kazumori Mise
mise-33@aist.go.jp
Hideomi Itoh
hideomi-itou@aist.go.jp
[1] Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[2] Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan
[3] National Institute of Advanced Industrial Science and Technology (AIST) Hokkaido, 2-17-2-1 Tsukisamu-higashi, Toyohira, Sapporo, Hokkaido 062-8517, Japan
[4] Niigata Agricultural Research Institute, 857 Nagakura-machi, Nagaoka, Niigata 940-0826, Japan

Introduction

Biological nitrogen fixation driven by diverse soil microorganisms is a distinct process providing the pedosphere with nitrogen, the major limiting factor for primary production [1]. Microbial players for nitrogen fixation (diazotrophs) in the soil have drawn significant attention since their discovery in the late nineteenth century [2]. In particular, the distribution and diversity of the diazotrophs in the soil have been one of the most active research topics and is constantly updated along with the accumulation of knowledge and technological innovations [3–5].

Although nitrogenase genes (*nif*) are conserved in a broad taxonomic range of prokaryotes [6], *nif* genes derived from *Alphaproteobacteria*, *Betaproteobacteria*, and *Cyanobacteria* have been frequently detected in various soil environments such as farmland, grassland, forests, rice paddy fields, riparian zones, and tundra by PCR amplicon surveys targeting *nif* genes [7–11]. Consequently, these bacteria are considered the primary nitrogen fixers in soil [12, 13]. In our previous work, however, shotgun metagenomic and metatranscriptomic analyses of paddy soil at one site in Japan have detected highly abundant *nif* genes and transcripts from the families *Anaeromyxobacteraceae* and *Geobacteraceae* within *Deltaproteobacteria* (also classified as the phyla *Myxococcota* and *Desulfobacterota,* respectively) compared with the conventional diazotrophic groups [14]. Several amplicon-based studies have also reported the occurrence of *Geobacteraceae* nitrogenase genes in soils [15–18]. Considering their prevalence across many soil types as revealed by 16S rRNA gene-based surveys [19–21], members of the families *Anaeromyxobacteraceae* and *Geobacteraceae* may thus represent universal and/or major components of diazotrophic microbiome in various terrestrial environments. However, these clades, which are well-known iron-reducing bacterial groups [22], have received considerably less attention as diazotrophs in soil than the conventional groups.

One potential problem is that genomic information of *Anaeromyxobacteraceae* and *Geobacteraceae* has been poorly represented in reference databases because pure isolates of these bacteria have been difficult to obtain. Fortunately, recent studies significantly enriched the reference sequence databases by isolating dozens of novel members within these families using our previously developed slurry incubation method; preincubated soil slurry and Reasoner's 2A (R2A) agar supplemented with fumarate were used as isolation source and medium, respectively, as described in details in Materials and methods section [23–31]. All of the novel isolates harbor nitrogenase genes, whereas some of them have been shown to present diazotrophic activities [25, 26, 29, 30].

Another problem is that amplicon sequencing tend to incur major biases in microbiome studies. Universal primers often fail to detect (even dominant) genes due to mismatches or abnormal GC contents of template DNA, whereas they may amplify homologous but unrelated genes [32–37]. Moreover, the primer sets (and accordingly PCR conditions) used for nitrogenase gene (typically *nifH*) amplification are not standardized among a plethora of amplicon sequencing studies [38]. This suggests that different studies bear different types and degrees of PCR biases. These call for extensive analyses of shotgun metagenomic data, rather than amplicon-based data, to establish minimally biased knowledge on diazotrophic communities in terrestrial microbiomes.

In this study, we aimed to re-evaluate the global distribution and diversity of the terrestrial diazotrophic microbiome considering the presence of *Anaeromyxobacteraceae* and *Geobacteraceae* bacteria. We scrutinized a global trend of the terrestrial diazotrophic microbiome using 1,451 shotgun metagenomic datasets, making full use of recently published genomic information on *Anaeromyxobacteraceae* and *Geobacteraceae* isolates. Our analyses revealed that *Anaeromyxobacteraceae* and *Geobacteraceae* are ubiquitous constituents of the diazotrophic microbiome in terrestrial ecosystems, particularly with high dominance in anaerobic environments.

Results and discussion

Diversity of *nif*-harboring genomes in public databases

We first reviewed the currently known diversity of nitrogen-fixing prokaryotes in public databases. KEGG included 7,152 bacterial genomes (KEGG ftp as of August 31, 2022), and 697 of them encoded all three structural genes of nitrogenase, namely *nifH*, *nifD*, and *nifK*. Among these genomes, those of *Alphaproteobacteria*, *Firmicutes*, and *Gammaproteobacteria*, as well as *Deltaproteobacteria*, were abundant (Fig. 1a).

The representations of *Geobacteraceae* and *Anaeromyxobacteraceae* sequences in public databases were recently improved. At the end of 2018, RefSeq contained 23 genomes from *Geobacteraceae* and 5 from *Anaeromyxobacteraceae*, whereas these numbers tripled by September 28, 2022. Approximately 50% of these increases could be attributed to isolates obtained by the slurry incubation method since 2019 (Table 1) [23–31], and all of these isolates bear the core *nif* genes (*nifHDK*) in their genomes. Apart from these isolates, we obtained two other distinct strains belonging to the genus *Geomonas*, namely Red32 (isolated from paddy soil in Joetsu, Niigata, Japan) and Red276 (pond sediment in Myoko, Niigata, Japan; Table 1). These strains displayed 96.3%–97.4% similarity (based on 16S rRNA gene sequences) to all *Geomonas* type strains, which was below the standard

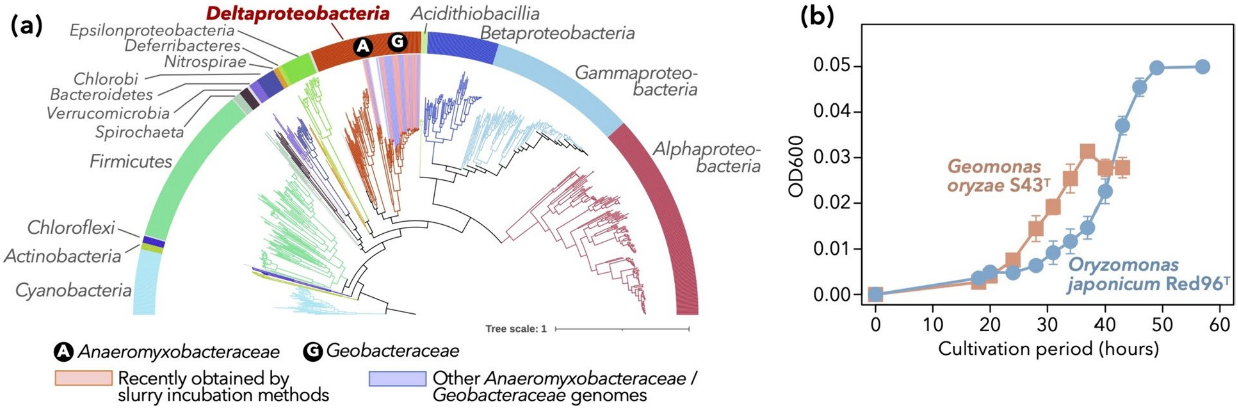

Fig. 1 Currently known diversity of diazotrophs. **a** A genome-based phylogenetic tree consisting of potential diazotrophic bacteria [i.e., the genomes of which harbor all three core genes of nitrogenase (*nifH*, *nifD*, and *nifK*)], including the genomes of new isolates of *Anaeromyxobacteraceae* and *Geobacteraceae* (Table 1). The colors of branches and the band surrounding the tree denote the phyla and proteobacterial classes. Genomes of families *Anaeromyxobacteraceae* and *Geobacteraceae*, the foci of the present study, are highlighted with circled letters (A and G) and colored backgrounds (blue and pink, respectively). **b** Growth curves of the type strains of two type species within the family *Geobacteraceae*, namely *Geomonas oryzae* S43^T and *Oryzomonas japonica* Red96^T. The two isolates were grown on MFM medium with N_2 as the sole nitrogen source. Average and standard deviation of each time point (n = 3) are indicated. Some error bars are shorter than the symbol size

threshold (98.65%) for species delineation [39]. The genomes of these strains also encoded *nifHDK*.

In addition to the presence of *nif* genes on the genomes, we confirmed the nitrogen-fixing activities of bacterial strains from these clades. In this study, we demonstrated that two type species within *Geobacteraceae*, namely *Geomonas oryzae* S43^T and *Oryzomonas japonica* Red96^T (Table 1) [27, 28], were able to grow on N_2 as the sole nitrogen source (Fig. 1b). While the acetylene reduction activity of some strains within *Geobacter* and *Geomonas* has been previously tested [25, 29, 40], the ammonium-independent steady growth of *Geomonas* and *Oryzomonas* suggests that nitrogen fixation is energetically available. The present result, combined with the previously reported N_2-dependent growth and acetylene reduction activity of *Anaeromyxobacter* and *Geobacter* strains [26, 40], indicate that *Anaeromyxobacteraceae* and *Geobacteraceae* are likely to be physiologically relevant to nitrogen fixation. We suspected that the use of their genomes as references would yield a better sensitivity in shotgun metagenomic analyses of deltaproteobacterial diazotrophs.

Global distribution of diazotrophs in terrestrial environments

To assess the global distribution of nitrogen-fixing populations, we collected 1,433 shotgun metagenomic datasets from public databases, namely NCBI SRA and MG-RAST [41, 42], coming from various environments including cropland soils, forest soils, grassland soils, paddy soils, sediments (including wetlands), and tundra soils (Fig. 2a and Table S1). Since metagenomes of Japanese soils (including volcanic soils) were poorly represented in public databases, we also collected 18 soil samples in Japan (Table S2) and sequenced their metagenomes. It should be noted that we only used metagenomic data in the present study, which indicate the quantity and diversity of potential diazotrophic microbes but do not serve as direct evidence of their diazotrophic activities.

Following preprocessing and curation of these datasets (i.e., merging of paired-end reads, quality filtering), we identified reads bearing nitrogenase genes (*nifDK* for molybdenum nitrogenase, *vnfDK* for vanadium nitrogenase, *anfDK* for iron-only nitrogenase [43]), 16S rRNA genes, or single-copy ribosomal protein genes conserved in most bacterial genomes [44] (Table S3). The phylogenetic compositions of 16S rRNA gene sequences indicated that 100 of these datasets could be contaminated by members of order *Lactobacillales* or plants' plastids (amounting up to 80.5% of *Lactobacillales* or 71.6% of Chloroplast 16S rRNA gene reads: Fig. S1), and the remaining 1,351 were used for further analyses (Fig. 2b). Some of the 1,351 metagenomes were redundant (e.g., technical replicates), so we clustered metagenomes taken from environmental samples within 1 km of each other. This ended up in 321 samples (Fig. 2b), each considered to bear independent information. The 321 samples were overall dominated by well-known soil-dwelling bacterial clades such as phyla *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, etc. (Fig. 2c) and showed no obvious hallmarks of abnormality or technical contamination.

Table 1 Novel bacterial members within the families *Anaeromyxobacteraceae* and *Geobacteraceae* that have been recently isolated using a slurry incubation method and published to date, including ones isolated in this study. Relevant publications, as well as genomic information, are also indicated

Strain	Accession No. of genomic data	Note	Reference
Anaeromyxobacteraceae			
Anaeromyxobacter diazotrophicus Red267^{T}	GCF_013340205.1		[24, 26]
Anaeromyxobacter oryzae Red232^{T}	GCF_023169945.1		[24]
Anaeromyxobacter paludicola Red630^{T}	GCF_023169965.1		[24]
Geobacteraceae			
Geomonas azotofigens Red51^{T}	GCF_018919395.1		[29]
Geomonas diazotrophica Red69^{T}	GCF_018919385.1		[29]
Geomesophilobacter sediminis Red875^{T} [a]	GCF_016458275.1		[31]
Geomonas propionica Red259^{T}	GCF_016458235.1		[31]
Geomonas anaerohicana Red421^{T}	GCF_016458305.1		[31]
Geomonas silvestris Red330^{T}	GCF_014193515.1		[23]
Geomonas paludis Red736^{T}	GCF_014193585.1	GCF_023221575.1 is included in KEGG	[23]
Geomonas paludis RG22	GCF_023221575.1		[30]
Geomonas limicola Red745^{T}	GCF_014193675.1		[23]
Oryzomonas japonica Red96^{T} [a]	GCF_008802365.1		[27]
Oryzomonas sagensis Red100^{T}	GCF_008802355.1		[27]
Oryzomonas rubra Red88^{T}	GCF_008369015.1		[27]
Geomonas oryzae S43^{T} [a]	GCF_004117875.1		[28]
Geomonas edaphica Red53^{T}	GCF_004917075.1		[28]
Geomonas ferrireducens S62^{T}	GCF_004917065.1		[28]
Geomonas terrae Red111^{T}	GCF_004791675.1		[28]
Geomonas fuzhouensis RG17^{T}	GCF_020179575.1		[30]
Geomonas agri RG53^{T}	GCF_020179605.1		[30]
Geomonas sp. Red32	JAKLOY010000000		This study
Geomonas sp. Red276	BLXW01000000		This study
Geomonas oryzisoli RG10^{T}	GCF_018986915.1	Included in KEGG	[25]
Geomonas subterranea RG2^{T}	GCF_019063845.1	Included in KEGG	[25]
Geomonas nitrogeniifigens RF4^{T}	GCF_019063885.1	Included in KEGG	[25]

[a] Type species of each genus; ICNP: International Code of Nomenclature of Prokaryotes

From each of the 321 samples, we detected 110–5,714,345 reads (median: 8,050 reads) of ribosomal protein genes listed in Table S3. The number of nitrogenase gene reads were normalized by the number of ribosomal protein gene reads, taking into account the differences in gene lengths between orthologs (corrected by RPKM: see Materials and Methods). The relative abundances of nitrogenase gene reads were higher in paddy soils, sediments, and tundra soils (i.e., anaerobic environments) than those in aerobic environments, namely cropland, forest, and grassland soils ($P < 0.05$ in post-hoc pairwise Brunner–Munzel test with Bonferroni correction, Fig. 2d; Please note that only four data belonged to tundra). On average, nitrogenase genes were detected 17.6 times more frequently in the anaerobic samples than aerobic samples. This is consistent with both the well-established notion that biological nitrogen fixation is an anaerobic process and the oxygen-sensitive nature of nitrogenase [45].

Relative abundances of nitrogenase gene reads exhibited major variations among samples from aerobic environments (i.e., cropland, forest, and grassland), with some harboring low numbers, and others dominated by diazotrophs. Although the reason for such variation is not clear and require further experimental validation, here we list several hypotheses. First, several soil physicochemical properties, including total carbon, nitrogen, and available phosphorus contents, have been shown to affect the abundance of species which encode nitrogenase gene [16, 46]. It is also possible that cropland and grassland samples are affected by the roots of leguminous plants and nodule symbionts therein; alphaproteobacterial N_2-fixing rhizobia such as *Bradyrhizobium*, *Azospirillum*, and *Mesorhizobium* were detected more frequently

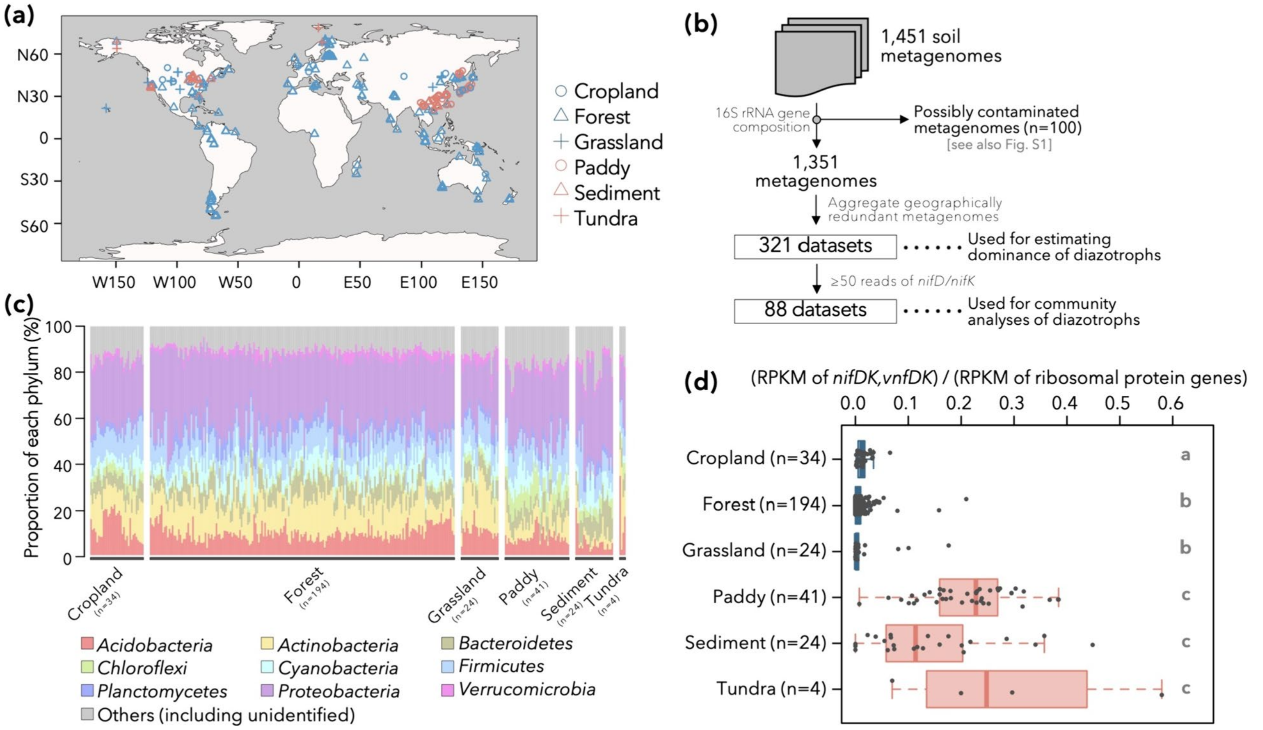

Fig. 2 Overview of metagenomic datasets used in this study and distribution of diazotrophs therein. **a** The sampling locations for each metagenomic datasets used in this study. Six types of environments are differentiated by the shapes and colors of symbols. **b** Filtering procedure of metagenomic datasets. The filtering criteria, as well as the aggregation of geographically similar samples, are explained in the panel. **c** Phylum-level prokaryotic community structure of the 321 metagenomic datasets estimated by 16S rRNA gene sequences. **d** The dominance of nitrogen-fixing population in each environment. For each of the 321 metagenomic datasets, the ratio of reads per kilobase of reference sequence per million sample reads (RPKM) of nitrogenase genes to the RPKM of ribosomal protein genes is displayed. The letters on the right side of the box indicates the statistical significance in RPKM ratio between different environmental categories ($P < 0.05$, Brunner–Munzel test with Bonferroni's correction). First, second, and third quantiles are indicated by solid lines. The whiskers, if any, denote 1.5*[interquartile range] from first or third quartile

in legume crop soils than in non-legume crop soils [46, 47]. Some samples from aerobic environments may be locally anaerobic, and this might explain the variance in the relative abundance of diazotrophs. We also acknowledge the ambiguity in distinguishing forest or grassland soils from wetland sediments. For example, samples from Disney Wilderness Preserve (DWP), which are labeled as "area of pastureland or hayfields" in MG-RAST and presented the highest relative abundances of nitrogenase genes among "grassland" samples, may have originated from wetland-like environments, as the landscape of DWP bears patches of wetlands [48].

Global diversity of diazotrophs in terrestrial environments

The taxonomic compositions of diazotrophic communities were further investigated for a more limited dataset of 88 samples, each of which comprised at least 50 sequences of *nifD/K* (Fig. 2b and Fig. S2). Please note that *nifD/K* serve as more accurate markers of diazotrophs compared with *nifH* (i.e., the conventional marker of diazotrophs [36]: see Materials and Methods for detail).

Reads encoding *nifD/K* from class *Deltaproteobacteria*, especially *Geobacteraceae* and *Anaeromyxobacteraceae*, were consistently dominant in anaerobic environments such as paddy soils and sediments, as well as in some of the aerobic samples (Fig. 3ab). While nitrogenase genes have a complicated history of horizontal gene transfer (HGT) that may hinder accurate taxonomic annotation [49, 50], Deltaproteobacterial NifD/K within Group I nitrogenase [49] are monophyletic (bootstrap value = 1.00, Fig. S3) and no clear hallmark of recent HGT [51]. Fortunately, a major part of deltaproteobacterial NifD/K within metagenomes belonged to this group (Fig. S4), and therefore it is unlikely that the dominance of *Deltaproteobacteria nifD/K* in this metagenomics is a byproduct of HGT. In addition, the proportion of deltaproteobacterial 16S rRNA genes (i.e., genes less prone to HGT) and frequency of deltaproteobacterial nitrogenase genes were significantly correlated (Fig. S5: Spearman's $\rho = 0.859$ and $P < 2.2\times10^{-16}$ when tested using all samples; $\rho = 0.738$ and $P < 2.2\times10^{-16}$ when tested using only anaerobic samples). These results suggest that members of

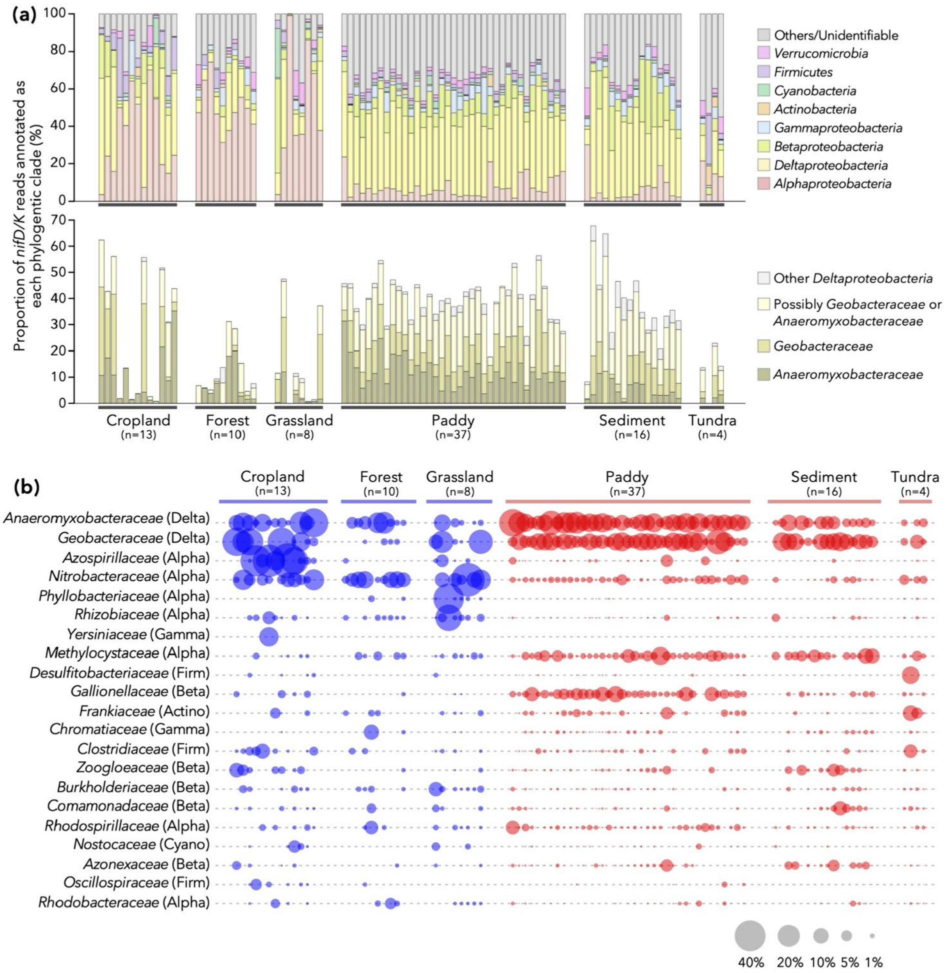

Fig. 3 Phylogenetic compositions of nitrogenase genes in the metagenomic datasets with at least 50 reads of *nifD* and *nifK* (n = 88 in total). **a** Upper panel: phylum- and proteobacterial class-level composition. Lower panel: breakdown of deltaproteobacterial composition at the family level. The category "Possibly *Geobacteraceae* or *Anaeromyxobacteraceae*" comprises deltaproteobacterial reads that were unannotated at the family level but received higher-level annotations consistent with family *Geobacteraceae* or family *Anaeromyxobacteraceae*. **b** Family-level distribution of *nifD* and *nifK* reads. The correspondence with the phylum- and proteobacterial class-level taxonomy is noted in parentheses: Delta, *Deltaproteobacteria*; Alpha, *Alphaproteobacteria*; Gamma, *Gammaproteobacteria*; Beta, *Betaproteobacteria*; Firm, *Firmicutes*; Actino, *Actinobacteria*; Cyano, *Cyanobacteria*. The area size (not the radius) of each plot is proportional to the relative abundance of each family within each dataset

Geobacteraceae and *Anaeromyxobacteraceae* are one of the prominent drivers of nitrogen fixation in terrestrial ecosystems. While previous studies in wheat-soybean rotation croplands [16] and paddy soils [14, 52] are in line with our results, the metagenomic datasets analyzed here covering a wide range of environments provide a generalizable insight into the potential contributions of these clades to nitrogen fixation processes in the pedosphere.

Other major clades within nitrogen-fixing populations included *Alphaproteobacteria*, such as *Nitrobacteraceae* and *Rhizobiaceae* (Fig. 3ab), although some (e.g., *Bradyrhizobium* and *Rhizobium*) in these families are symbiotic diazotrophs and thus possibly incapable of independent nitrogen fixation outside their host plants. The community compositions were significantly different between the aerobic and anaerobic samples (permutational analysis of variance (PERMANOVA) of UniFrac distances, $R^2 = 0.114$, $P = 0.001$), as further evidenced by a distinct grouping of the two types of samples in nonmetric multidimensional scaling (NMDS) analysis (Fig. S6).

Another characteristic of the diazotrophic communities of anaerobic environments (with the exception of tundra) is the high similarity between samples (Fig. 3). While all the paddy soil and sediment samples are dominated by *Deltaproteobacteria* and present overall low beta-diversity levels, environments such as cropland, forest, and grassland are dominated by more diverse clades of diazotrophs showing high beta-diversity levels. One possible explanation is the heterogeneity among aerobic samples: some of the cropland, forest, and grassland samples may be associated with leguminous vegetations (i.e., affected by nodule-associated bacteria) [18, 46] or originated from soil physicochemical properties/conditions [16], but they are not explicitly considered in this study. Another explanation for this divergence in community structures is ecological drift [53, 54]. Diazotrophs have smaller population sizes in aerobic samples (Fig. 2d); thus, their communities are expected to be more sensitive to ecological drift, resulting in increased beta-diversities between communities as previously shown [55].

Although we used a limited dataset of 88 samples bearing at least 50 *nifDK* sequences for the taxonomic composition analysis, the bias introduced by this manipulation is unlikely to be critical. First, the selected samples do not necessarily present high relative abundances of diazotrophs, since the number of total reads greatly varies between samples (Fig. S2ab). Second, the abundance ratio of nitrogenase genes to ribosomal protein genes explain only 5.0% the phylogenetic diversity of *nifDK* ($R^2 = 0.050$, $P = 0.036$) within aerobic samples (Fig. S2c).

As a side note, we also analyzed the abundance and phylogeny of *nifH*, a conventional marker for nitrogen-fixing populations. Because the lengths of *nifH*, *nifD*, and *nifK* genes are approximately 3:5:5 (Table S3), the number of these genes should also be around 3:5:5 in each metagenome. However, some samples harbored disproportionally higher number of *nifH* compared with *nifD/K*: nine of the metagenomic samples included 1.5 times or higher number of *nifH* reads than can be expected from the number of *nifD/K* reads (blue points in Fig. S7a). This implies that some soil samples bear significant amount of pseudo-*nifH* genes [36], which are encoded on prokaryotic genomes lacking other essential components of nitrogenase genes (e.g., *nifD* and *nifK*). We suspect that *nifD/K*, rather than *nifH*, serve as a reliable marker gene for nitrogen-fixing populations (especially in shotgun metagenomic studies). Regarding 79 samples with lower amounts of pseudo-*nifH*, deltaproteobacterial *nifH* were dominant (Fig. S7b) in congruence with the results of *nifDK* analyses (Fig. 3a).

An approximate estimation of the global dominance of deltaproteobacterial diazotrophs

Analyses of the global metagenomic dataset indicated that anaerobic environments harbor high abundances of diazotrophic prokaryotes and that *Anaeromyxobacteraceae* and *Geobacteraceae* are the dominant diazotrophs in these environments. Wetlands (possibly including waterlogged paddy soils) represent between 5.2% [56] and 8% [57] of all lands, with microbial biomass carbon therein amounting to 10.3% of the microbial biomass in all lands [calculated from the data presented in [56]]. Thus, although wetland is a limited area of land, given that the relative abundance of nitrogen fixers was 17.6 times higher in anaerobic microbial communities than in aerobic ones (Fig. 2d), wetland could be a large reservoir of nitrogen fixers on terrestrial environments.

Biases behind amplicon sequencing of nitrogenase genes

The prevalence of diazotrophic *Geobacteraceae* has actually been reported in some of PCR amplicon sequencing analyses of *nif* genes [15, 16, 58], but the dominance of *Anaeromyxobacteraceae* has been overlooked in such PCR-based analyses. We suspected that this discrepancy is due to PCR biases behind amplicon sequencing. It is commonly accepted that results of amplicon sequencing are dependent on a series of PCR conditions such as primer sets and DNA polymerases [37, 59]. The GC contents of templates, as well as primer mismatches, can also affect the amplification efficiency and therefore cause biases [34, 35, 37]. Notably, *nif* genes of *Anaeromyxobacteraceae* have higher GC contents (65.6–69.7%) than those of the other bacteria (Fig. S8).

To elucidate the PCR biases of *nif* genes, we performed amplicon sequencing of nitrogenase genes in six soil DNA samples and compared the results directly with shotgun sequencing of the same samples. We prepared amplicon libraries under ten PCR conditions with different primer sets and DNA polymerases (Tables S4 and S5). Please note that we here targeted *nifH*, rather than *nifDK*, for the sake of consistency with conventional amplicon sequencing methods. Primer mismatches will not be extensively discussed here, because *nifH* of *Anaeromyxobacteraceae* and other clades present similar identities to the primers (Fig. S9).

As expected, we found that the phylogenetic compositions of *nifH* amplicons were dependent on type of DNA polymerases and primer sets (Fig. S10a–f), and the discrepancy was particularly remarkable in the proportion of *Anaeromyxobacteraceae nifH*. *Anaeromyxobacteraceae nifH* were consistently more highly represented in KOD One libraries than in DreamTaq libraries (Fig. S10g). In addition, their proportion in shotgun metagenomic sequences were comparable to those in KOD One libraries, although dependent on sample identities and primer sets (Fig. S10h). These suggest that DreamTaq failed to amplify *Anaeromyxobacteraceae nifH*. What we focus on here is the high GC contents of *nif* genes in *Anaeromyxobacteraceae* (Fig. S8). According to the manufacturers' reports (https://lifescience.toyobo.co.jp/user_data/pdf/products/manual/KMM-101_201.pdf [in Japanese; accessed Jan 5, 2024]), KOD One is robust to amplify GC-rich templates, while DreamTaq shows a low performance (https://www.thermofisher.com/order/catalog/product/EP1701?SID=srch-srp-EP1701 [accessed Jan 5, 2024]), which aligns with the present results. We speculate that GC richness of *Anaeromyxobacteraceae nifH* may be one reason why they have been poorly represented in amplicon surveys.

We also argue that these results represent the major biases behind amplicon sequencing. Provided that *nifH* gene compositions were largely dependent on the type of DNA polymerases and primer sets, comparing the results obtained from multiple studies might be difficult. In this respect, meta-analysis of shotgun metagenomic data should be a straightforward, solid, and less biased approach (as has been discussed in Kim et al. [60]).

Even in shotgun metagenomic sequencing, it should be noted that the abundance of diazotrophic *Anaeromyxobacteraceae* may be underestimated. First, library preparation for shotgun sequencing often involves PCR (typically 8–12 cycles), which may fail to amplify GC-rich nucleotide fragments [61, 62]. Second, Illumina sequencing technology is known to be biased against sequencing GC-rich nucleotide fragments even in shotgun sequencing [63]. Considering these biases, the proportion of *Anaeromyxobacteraceae* in the soil, the nitrogenase genes of which are GC-rich (66.9%–69.0%, 65.6%–67.0%, and 66.9%–69.7% for *nifH*, *nifD*, and *nifK*, respectively, Fig. S8), might be even higher than estimated in this study. The former issue may be addressed using PCR-free library preparation protocols [62]. Long-read sequencers (i.e., PacBio and Nanopore) are less prone to GC bias [63] and potentially rectify the latter issue, although their current yield is orders of magnitude smaller than those of short-read sequencers such as Illumina HiSeq and NovaSeq, and thus currently not a good fit for the characterization of samples as heterogenous and rich in diversity as soil metagenomes.

Benefits of expanding culture collection

Previous and current efforts to enrich culture collections [23, 24, 26–29, 31] have substantially expanded the available repertoire of *Anaeromyxobacteraceae* and *Geobacteraceae* strains. In fact, an average of 56.9% and 23.9% of NifD/K sequences derived from *Anaeromyxobacteraceae* and *Geobacteraceae* members, respectively, displayed higher similarity to our novel strains than to any other nitrogenase sequence in KEGG from these families (Fig. 4).

Interestingly, this trend was consistent among a wide variety of environments including aerobic and anaerobic environments, although the majority of the novel strains were isolated from paddy soils or sediments under anaerobic conditions. Based on the present and previous findings, paddy soils and sediments appear to be promising environments for isolating free-living diazotrophs, representing diverse terrestrial environments including aerobic environments such as cropland, forests and grassland.

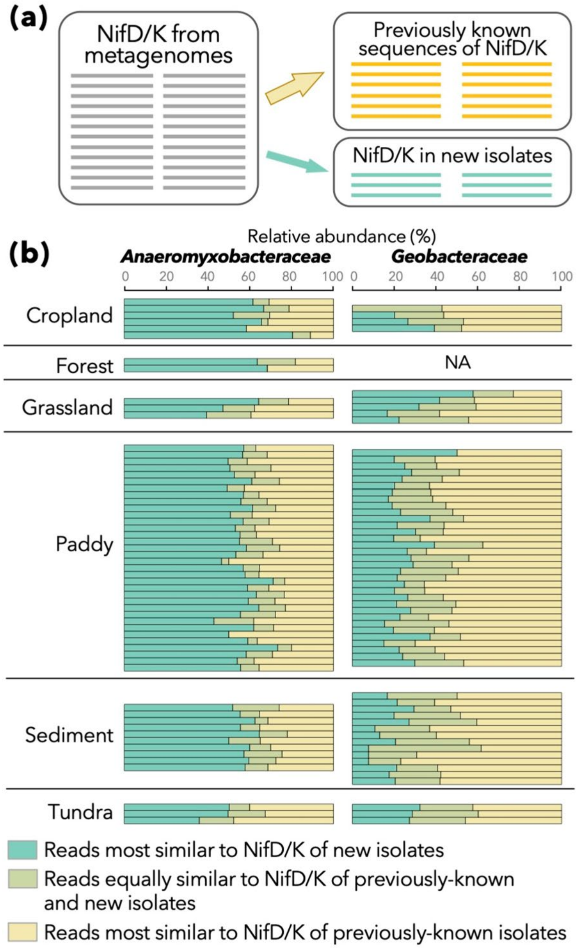

Fig. 4 Contribution of nitrogenase gene sequences from newly isolated strains in the bioinformatic analyses of metagenomes. **a** A schematic of the analysis. NifD/K sequences annotated as *Anaeromyxobacteraceae* or *Geobacteraceae* in metagenomes were mapped onto already known sequences of NifD/K (right-upper) and those in our new isolates (right-bottom). Only the top hit for each query sequence (i.e., one from metagenomes) was considered. **b** Relative abundance of metagenome-derived NifD/K sequences that were most similar to already known sequences (yellow) and those from our new isolates (green), as well as those equally similar to the nitrogenase genes of already known genomes and our new isolates (dim green), are summarized. Only datasets with 10 or more sequences of NifD/K for each family are displayed

Conclusions and outlook

Contrary to the conventional view, our large-scale comparative metagenomics analyses revealed the global distribution and substantial abundance of *Anaeromyxobacteraceae* and *Geobacteraceae* in terrestrial diazotrophic microbiome, highlighting the potential importance of *Deltaproteobacteria* members (phyla *Myxococcota* and *Desulfobacterota*) in terrestrial, especially anaerobic, ecosystems. Although *Anaeromyxobacteraceae* and *Geobacteraceae* have been well known as iron- and other metals- reducing bacteria in soil environments, this study is the first to report that they are the most dominant group of terrestrial diazotrophic microbiome on a global scale.

Moreover, nitrogen-fixing bacteria have long been considered useful microorganisms for improving soil nitrogen fertility, and methods to promote their activity have been developed for sustainable agriculture [64]. For example, in paddy soils, recent studies showed that application of iron-bearing materials could enhance the nitrogen-fixing activities of indigenous iron-reducing bacteria within the families *Anaeromyxobacteraceae* and *Geobacteraceae* and maintain rice yields under reduced nitrogen-fertilizer application [65, 66]. Given the ubiquity of iron-reducing diazotrophs (Fig. 3), this strategy may be effective in a variety of other crop fields. More generally, careful and precise updates of our understanding of functional microorganisms in soil environments should advance such attempts towards sustainable agriculture.

It should be noted that the pivotal thing for microbiome discovery is to improve the accuracy of metagenomics, i.e., to expand the available genomic information of microorganisms. In this study, thousands of obtained nitrogenase sequences exhibited high proximity to our newly isolated strains. Our results warrant further efforts to improve culture collections, which would fill the knowledge gaps in the diversity and ecology of diazotrophs. Especially in soil environments, the enormity of uncultured but predominant clades of prokaryotes, as represented by members of *Acidobacteria* and *Verrucomicrobia* [67–69], is widely recognized. To advance our knowledge of the terrestrial diazotrophic microbiome, strategies for their cultivation and isolation should be also updated, for example, by using single-cell sorting.

There is no doubt that *Anaeromyxobacteraceae* and *Geobacteraceae* are important diazotrophic members in soils that should not be underestimated or undervalued as they have been. However, unfortunately, it is impossible to estimate how much *Anaeromyxobacteraceae* and *Geobacteraceae* actually contribute to nitrogen fixation in soil environments based on the results of this study alone, since the contribution cannot be directly inferred from the detected amount of genes. The insights into the contribution of each diazotrophic taxon to terrestrial nitrogen fixation will be foreseeable, for instance, through stable isotope probing (SIP) with $^{15}N_2$ under more natural conditions, using various soil samples. Although preliminary, a recent study based on ^{15}N-DNA-SIP analysis revealed a high contribution of *Anaeromyxobacteraceae* and *Geobacteraceae* for nitrogen fixation in a paddy soil [70], supporting our conclusion.

Materials and methods

Isolation and genomic sequencing of new soil strains

The *Geomonas* strains Red32 and Red276 were isolated from paddy soil (Joetsu, Niigata, Japan) and pond sediment (Myoko, Niigata, Japan) following the slurry incubation method used to isolate new members of *Geobacteraceae* [27, 28]. The soils collected from the paddy field in Nagaoka, Niigata, Japan were air-dried, placed in a 15-mL serum bottle and suspended in distilled water (soil:water, 2:3, *w*/*v*). After autoclaving at 120°C for 20 min, 0.1 g of undried soil was added to the bottle as a microbial inoculum with and without vitamin solution for strains Red276 and Red32, respectively [26]. Then, we sealed the bottles with butyl rubber stoppers and aluminum caps, replaced headspace gas with with N_2/CO_2 (80:20, *v*/*v*), and incubated them at 30°C for 2 weeks without shaking. Afterward, 200 µL of incubated soil slurry was transferred to a new bottle of autoclaved soil slurry and incubated at 30 °C for 2 weeks. After repeating this step once (for strain Red276) or twice (for strain Red32), the incubated soil slurry was streaked on 1.5% agar plates of the R2A broth "DAIGO" (Nihon Pharmaceutical, Tokyo, Japan) supplemented with 5 mM disodium fumarate. The plates were incubated at 30°C for 10 days under anaerobic conditions using the AnaeroPack system (Mitsubishi Gas Chemical, Tokyo, Japan). Red-colored colonies, a typical hallmark of *Geobacteraceae* strains [23, 27, 28, 71, 72], were purified by a single-colony isolation using the same medium plates. Genomic DNA was extracted from the two isolated strains using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) and sequenced using an Illumina HiSeq sequencer (Illumina, CA, USA) for 2×150 paired-end configuration. The resulting sequences were assembled using Velvet v1.2.10 [73] as previously described [27, 28].

Diazotrophic activity assay

Following a 5-day culture in nitrogen-free modified freshwater medium (MFM) as previously described [27, 28], the cells of *Geomonas oryzae* S43^T and *Oryzomonas japonica* Red96^T were transferred to serum bottles containing 20 mL of nitrogen-free MFM [28] and headspace gas was replaced with N_2 gas. No contamination of ammonia in used N_2 gas was confirmed by no growth of non-diazotrophic *Anaeromyxobacter* strain, *A. dehalogenans* 2CP-1^T [26]. Bacterial growth was monitored by measuring the suspension absorbance using a spectrophotometer (UV-1900 UV-visible spectrophotometer, Shimadzu, Kyoto, Japan) at a wavelength of 600 nm. The experiments were performed in triplicate.

Preparation of custom database

We used KEGG database (as of August 31, 2022) for functional gene annotations [74]. To increase the sensitivity for genes from *Anaeromyxobacteraceae* and *Geobacteraceae*, we customized KEGG database by adding genomes belonging to these families obtained using slurry incubation methods (Table 1). Genomes already included in KEGG were not added. We predicted their coding sequences (CDS) using Prodigal [75] with default parameters, annotated them using KofamScan version 1.3.0 and KOfam version 2022-08-01 with default parameters [76], and concatenated the CDS with KEGG database (including those received no K number). The phylogenies of NifD and NifK within Group I [49] were determined using MAFFT v7.505 (with "--auto" option) and FastTree 2.1.11 [77, 78] with a bootstrap test of 100 iterations (otherwise default parameter settings). For the bootstrapping tests, we also used "CompareToBootstrap.pl" script (http://www.microbesonline.org/fasttree/treecmp.html, accessed April 18, 2023) to merge the resampled trees.

Phylogenetic analysis of bacterial genomes harboring *nif* genes

From the aforementioned custom database, we screened genomes harboring a set of *nif* core genes, namely *nifH* (K02588 in KEGG), *nifD* (K02586), and *nifK* (K02591). Archaeal genomes were excluded from the analysis. The universal single-copy gene sequences were identified from each genome, translated amino acid sequences, and mapped onto multiple sequence alignment (MSA) of GTDB R207 using GTDB-Tk v2.1.0 [44, 75, 79, 80]. Here "identify" and "align" commands were used with default parameter settings. The MSA was fed into FastTree (default parameters) and a phylogenetic tree was constructed. The tree was manually rerooted using *Cyanobacteria* as the outgroup [81] and visualized on the iTOL server [82].

GC content of 16S rRNA genes and nitrogenase genes among bacterial genomes

Ribosomal RNA genes were identified from each of the bacterial genomes in the custom database explained above using barrnap version 0.9 (https://github.com/tseemann/barrnap; accessed April 18, 2023). Only 16S rRNA sequences with 1000 bases or longer were picked. For each genome with at least one valid 16S rRNA gene sequence and all of the identified *nifH*, *nifD* and *nifK* (identified as previously described), the GC contents of 16S rRNA genes, *nifH*, *nifD*, and *nifK* were calculated. When a genome had multiple copies of each gene, GC content was calculated for the concatenated sequence of these copies. Any ambiguous base was excluded from the calculation of GC content.

Soil collection and shotgun metagenomic sequencing

We collected 18 surface soil samples from various agricultural fields in Japan at an approximate depth of 0–5 or 0–10 cm (Table S1). Following the removal of plant residues and additional water from the surface, the soil samples were stored at −80 °C or −30 °C until further use for DNA extraction. Soil DNA was extracted from 0.5 g (wet weight) of each soil sample using the ISOIL for Beads Beating Kit (Nippon Gene, Tokyo, Japan) according to the manufacturer's instruction with the following modifications: prior to the beads beating step, 0.02 g skim milk was added to the lysis buffer to improve the extraction efficiency [83] and post-elution purification using RNase A (Takara, Shiga, Japan) and DNA Clean & Concentrator (Zymo Research) according to the manufacturer's introduction. Purified DNA was quantified using Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) with Qubit dsDNA HS Assay Kits (Invitrogen). The construction of DNA libraries, shotgun sequencing on an Illumina MiSeq sequencer, and merging of paired-end sequences were performed as described previously [14]. Regarding the other 6 soil samples, DNA was extracted from 0.25 g of each soil using DNesay PowerSoil Pro Kits (QIAGEN, Hulsterweg, Netherland) following the manufacturer's instruction. Shotgun sequencing library was prepared using MGIEasy FS DNA Library Prep (MGI Tech, Guangdong, China), where the duration of fragmentation reaction was customized to four minutes and library amplification was performed for eight cycles. MGIEasy Circularization Kit and DNBSEQ-G400RS High-throughput Sequencing Kit Set were used to construct DNBs, which were sequenced on DNBSEQ-G400 (MGI Tech) under 2x200 bp paired-end mode. Soil pH(H_2O) and electrical conductivity were measured in a suspension sample with soil-water ratio of 1:5 (w/w). Soil total carbon and nitrogen contents were determined using dry combustion method. Crop types and chemical properties of Japanese soils used in this study were summarized in Table S2.

Collection of publicly available metagenomic data and their quality assessment

We further collected reusable datasets of bulk soil metagenomes on INSDC [84] and MG-RAST [85] that met the following criteria: (i) derived from outdoor samples exempted from post-sampling treatments that can affect the microbial community structure; (ii) sequenced on Illumina MiSeq, HiSeq, MiniSeq, NextSeq or NovaSeq (i.e., state-of-the-art, highly accurate sequencers); and

(iii) reported in the peer-reviewed literature (with the exception of data obtained by the National Ecological Observatory Network). Moreover, the datasets from rhizosphere soils were not used in this study because they are extensively and dynamically affected by the plant roots [86] and not representative of the soil microbial communities. In total, we collected 1451 datasets as listed in Table S1 [47, 87–121]. The latitude and longitude of each sampling site were obtained from public databases [INSDC BioSamples database [122] and MG-RAST] and verified with the descriptions in each publication. The INSDC data were directly obtained from DDBJ server, whereas those on MG-RAST were fetched using MG-RAST API (with the option "file=050.1").

The collected metagenomic data underwent extensive curations, followed by homology searches to detect nitrogenase genes, ribosomal protein genes, and 16S rRNA genes. Detailed procedures are provided in the supplementary information. After a series of data curation, we decided to use 1,333 metagenomes from public databases and newly sequenced 18 metagenomes for downstream analyses. Some of the metagenomes were geographically redundant, so we merged metagenomes from samples taken within < 1 km and treated them as one sample. The distances between sampling locations were calculated based on the latitude and longitude of each sample using the geodesic module in GeoPy (https://geopy.readthedocs.io/en/stable/#; accessed Jan 5, 2024).

Gene annotations of metagenomic reads

To determine the nitrogen-fixing populations within each metagenomic dataset, the filtered sequences were subjected to homology search against the custom database explained above (i.e., KEGG database supplemented with *Anaeromyxobacteraceae* and *Geobacteraceae* genomes), followed by the taxonomic annotation of nitrogenase gene reads. In short, we determined the relative abundance of nitrogenase-harboring prokaryotes and their taxonomic composition for each sample.

Although the details are explained in the supplemental text, here we note three key strategies. First, we used *nifD*, *nifK*, *vnfD*, *vnfK*, *anfD*, and *anfK* as the marker genes. *nifH* was not used for this purpose because the partial primary structure of NifH can be confused with those of other proteins irrelevant to nitrogen fixation [36]. Second, we normalized the number of reads by those of single-copy prokaryotic ribosomal protein genes [81], rather than by the total number of metagenomic reads that may be affected by plant- and animal-derived sequences. Third, we used phylogenetic placement, rather than a simple homology search, for taxonomic annotation of *nif* gene reads. The reliability of each taxonomic annotation was calculated based on the likelihood of phylogenetic relationships between metagenomic reads and reference sequences, and therefore we were able to abandon uncertain annotations. For example, short fragmented reads may bear little phylogenetic signals, and annotations of such reads were to be unreliable and discarded.

Beta-diversity analyses

Beta-diversity between any pair of diazotrophic communities was calculated using the average of UniFrac distances for NifD and NifK, which were determined based on the results of phylogenetic placement. We used NMDS with two dimensions to summarize overall beta-diversity between communities. We also performed PERMANOVA with 999 times permutation to test the null hypothesis that community structures of diazotrophs are similar between aerobic and anaerobic environments.

Homology analyses between NifD/K of metagenomes and isolate genomes

We further mapped the NifD/K sequences annotated as family *Anaeromyxobacteraceae* or *Geobacteraceae* in metagenomic reads to NifD/K sequences from that family in our custom database using the Needleman–Wunsch algorithm implemented in USEARCH v11.0.667 [123]. We obtained the sequence similarity between each read and its nearest sequence in the database. We counted the number of reads for which the nearest sequence is from the genomes of bacterial isolates obtained via the slurry incubation method (Table 1).

Amplicon sequencing of nitrogenase genes using popular primer sets

We compared the results of shotgun metagenomic sequencing and amplicon sequencing of nitrogenase genes using six of the Japanese soil samples. Using the four pairs of universal primers (Table S4) and three DNA polymerases that differ in performance (DreamTaq DNA Polymerase [ThermoFisher Scientific], Ex Taq Hot Start Version [Takara], and KOD One [TOYOBO]), and we amplified *nifH* genes contained in each soil metagenome. Here we performed two-step tailed PCR to construct Illumina library, consisting of the first PCR to amplify *nifH* genes and the second PCR to attach index sequences to the amplicons. The amplicon of first PCR were cleaned up using AMPure XP (Beckman Coulter, Brea, CA, USA) before subjected to the second round of PCR. Detailed PCR conditions are summarized in Tables S4 and S5. The final PCR products were electrophoresed on agarose gels, purified using Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA), and sequenced on Illumina iSeq in a paired-end mode (151 bp × 2). For each combination of soil samples, primer pairs, and DNA polymerases, we amplified *nifH* genes and sequenced them in triplicates. The obtained reads underwent error correction using DADA2 [124], and the amplicon sequence variants (ASVs) were further filtered to eliminate chimeras and non-specific amplicons. The filtered ASVs were taxonomically annotated using phylogenetic placement. Details are explained in supplemental method.

Estimation of mismatches between *nif*-harboring genomes and *nifH* universal primers

To determine the mismatches between prokaroytic nitrogenase genes and their universal primers, we mapped sequences of seven primers (PolF, PolR, nifH-F, nifH-R, Ueda19F, Ueda407R, and univ463r: Table S4) onto 12 prokaryotic *nifH* sequences (*Anaeromyxobacter* sp. Fw109-5, *Anaeromyxobacter* sp. K, *A. diazotrophicus* Red267^T, *A. oryzae* Red232^T, *A. paludicola* Red630^T, *Azospirillum brasilense* Sp7^T, *Azotobacter vinelandii* DJ, *Bradyrhizobium diazoefficiens* USDA110^T, *Clostridium acetobutylicum* ATCC824^T, *Frankia casuarinae* CcI3^T, *Geomonas oryzae* S43^T, and *Oryzomonas japonica* Red96^T). We referred to annotations on KEGG or NCBI RefSeq to collect *nifH* sequences. In cases where one genome owned multiple copies of *nifH*, we selected one copy that was accompanied by *nifDK* in their neighborhood [36]. The collected *nifH* genes were aligned using MAFFT v7.505 (with "--auto" option), and then the primer sequences were manually aligned onto the MSA.

Throughout the study, we used SeqKit v0.16.1/v2.2.0 [125] and R 4.0.5/4.1.1 [126], including the package "vegan" [127], to handle fastq and fasta files and to perform statistical tests, respectively.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-024-01812-1.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

Part of the metagenomic data used in this study was provided by the National Ecological Observatory Network (NEON) via MG-RAST. We thank Haruka Ooi and Emiko Kobayashi (National Institute of Advanced Industrial Science and Technology) for literature survey, Yumi Sugisawa (National Institute of Advanced Industrial Science and Technology) and Kawata Laboratory (Tsukuba, Ibaraki, Japan) for performing experiments, and anonymous farmers for providing soil samples. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics and the SHIROKANE supercomputer at Human Genome Center, The Institute of Medical Science, The University of Tokyo.

Authors' contributions

Y.M., K.M., K.S., and H.I. designed the study and supervised the project. Y.M., Z.X., and Z.Z. performed genomic sequencing and diazotrophy assays of bacterial strains. K.M., Y.S., and H.I. collected Japanese soil samples, and Y.M. and K.M. performed metagenomic sequencing of these samples. Y.M. and H.I. performed the primary bioinformatic analysis of the metagenomic dataset. K.M. curated and analyzed genomic and metagenomic datasets, and performed amplicon sequencing analysis. Y.M., K.M., and H.I. wrote the paper with substantial input from all authors.

Funding

This work was financially supported by the CANON Foundation, JSPS KAKENHI Grant Numbers JP20H00409, JP20H05679, JP20K15423, and JP22K18029, MEXT KAKENHI Grant Number JP22H04894, JST-Mirai Program grant JPMJMI20E5, and JPNP18016 commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Availability of data and materials

Genomic sequences obtained in this study have been deposited in GenBank. Shotgun metagenomic and amplicon sequences have been deposited in DDBJ DRA. See Tables 1, S1, and S6 for accession numbers.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 28 April 2023 Accepted: 9 April 2024
Published online: 24 May 2024

References

1. Vitousek PM, Cassman K, Cleveland C, Crews T, Field CB, Grimm NB, et al. Towards an ecological understanding of biological nitrogen fixation. Biogeochemistry. 2002;57:1–45. https://doi.org/10.1023/A:1015798428743.
2. Beijerinck MW. Die bacterien der papilionaceenknöllchen. Botanische Zeitung. 1888;46:725–35.
3. Bürgmann H, Widmer F, Von Sigler W, Zeyer J. New Molecular Screening Tools for Analysis of Free-Living Diazotrophs in Soil. Appl Environ Microbiol. 2004;70:240–7. https://doi.org/10.1128/AEM.70.1.240-247.2004.
4. Hsu S-F, Buckley DH. Evidence for the functional significance of diazotroph community structure in soil. The ISME Journal. 2009;3:124–36. https://doi.org/10.1038/ismej.2008.82.
5. Nelson MB, Martiny AC, Martiny JBH. Global biogeography of microbial nitrogen-cycling traits in soil. Proc Natl Acad Sci. 2016;113:8033–40. https://doi.org/10.1073/pnas.1601070113.
6. Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. BMC Genomics. 2012;13:162. https://doi.org/10.1186/1471-2164-13-162.
7. Che R, Deng Y, Wang F, Wang W, Xu Z, Hao Y, et al. Autotrophic and symbiotic diazotrophs dominate nitrogen-fixing communities in Tibetan grassland soils. Sci Total Environ. 2018;639:997–1006. https://doi.org/10.1016/j.scitotenv.2018.05.238.
8. Gaby JC, Buckley DH. A global census of nitrogenase diversity. Environ Microbiol. 2011;13:1790–9. https://doi.org/10.1111/j.1462-2920.2011.02488.x.
9. Wang Q, Quensen JF, Fish JA, Kwon Lee T, Sun Y, Tiedje JM, et al. Ecological Patterns of *nifH* Genes in Four Terrestrial Climatic Zones Explored with Targeted Metagenomics Using FrameBot, a New Informatics Tool. mBio. 2013;4. https://doi.org/10.1128/mBio.00592-13.
10. Yu Y, Zhang J, Petropoulos E, Baluja MQ, Zhu C, Zhu J, et al. Divergent Responses of the Diazotrophic Microbiome to Elevated CO_2 in Two Rice

Cultivars. Front Microbiol. 2018;9 https://doi.org/10.3389/fmicb.2018.01139.

11. Zhu C, Friman V, Li L, Xu Q, Guo J, Guo S, et al. Meta-analysis of diazotrophic signatures across terrestrial ecosystems at the continental scale. Environ Microbiol. 2022;24:2013–28. https://doi.org/10.1111/1462-2920.15984.
12. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nat Rev Microbiol. 2018;16:263–76. https://doi.org/10.1038/nrmicro.2018.9.
13. Mahmud K, Makaju S, Ibrahim R, Missaoui A. Current Progress in Nitrogen Fixing Plants and Microbiome Research. Plants. 2020;9:97. https://doi.org/10.3390/plants9010097.
14. Masuda Y, Itoh H, Shiratori Y, Isobe K, Otsuka S, Senoo K. Predominant but previously-overlooked prokaryotic drivers of reductive nitrogen transformation in paddy soils, revealed by metatranscriptomics. Microbes Environ. 2017;32:180–3. https://doi.org/10.1264/jsme2.ME16179.
15. Calderoli PA, Collavino MM, Behrends Kraemer F, Morrás HJM, Aguilar OM. Analysis of *nifH*-RNA reveals phylotypes related to *Geobacter* and *Cyanobacteria* as important functional components of the N_2-fixing community depending on depth and agricultural use of soil. MicrobiologyOpen. 2017;6. https://doi.org/10.1002/mbo3.502.
16. Fan K, Delgado-Baquerizo M, Guo X, Wang D, Wu Y, Zhu M, et al. Suppressed N fixation and diazotrophs after four decades of fertilization. Microbiome. 2019;7:143. https://doi.org/10.1186/s40168-019-0757-8.
17. Feng M, Adams JM, Fan K, Shi Y, Sun R, Wang D, et al. Long-term fertilization influences community assembly processes of soil diazotrophs. Soil Biol Biochem. 2018;126:151–8. https://doi.org/10.1016/j.soilbio.2018.08.021.
18. Wang C, Zheng MM, Chen J, Shen RF. Land-use change has a greater effect on soil diazotrophic community structure than the plant rhizosphere in acidic ferralsols in southern China. Plant Soil. 2021a;462:445–58. https://doi.org/10.1007/s11104-021-04883-3.
19. Mitter EK, Germida JJ, de Freitas JR. Impact of diesel and biodiesel contamination on soil microbial community activity and structure. Sci Rep. 2021;11:10856. https://doi.org/10.1038/s41598-021-89637-y.
20. Pecher WT, Martínez FL, DasSarma P, Guzmán D, DasSarma S. 16S rRNA Gene Diversity in Ancient Gray and Pink Salt from San Simón Salt Mines in Tarija, Bolivia. Microbiology Resource Announcements. 2020;9:e00820–0. https://doi.org/10.1128/MRA.00820-20.
21. Sun W, Xiao E, Pu Z, Krumins V, Dong Y, Li B, et al. Paddy soil microbial communities driven by environment- and microbe-microbe interactions: A case study of elevation-resolved microbial communities in a rice terrace. Sci Total Environ. 2018;612:884–93. https://doi.org/10.1016/j.scitotenv.2017.08.275.
22. Weber KA, Achenbach LA, Coates JD. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. Nat Rev Microbiol. 2006;4:752–64. https://doi.org/10.1038/nrmicro1490.
23. Itoh H, Xu Z, Masuda Y, Ushijima N, Hayakawa C, Shiratori Y, et al. *Geomonas silvestris* sp. nov., *Geomonas paludis* sp. nov. and *Geomonas limicola* sp. nov., isolated from terrestrial environments, and emended description of the genus *Geomonas*. Int J Syst Evol Microbiol. 2021;71:004607. https://doi.org/10.1099/ijsem.0.004607.
24. Itoh H, Xu Z, Mise K, Masuda Y, Ushijima N, Hayakawa C, et al. *Anaeromyxobacter oryzae* sp. nov., *Anaeromyxobacter diazotrophicus* sp. nov. and *Anaeromyxobacter paludicola* sp. nov., isolated from paddy soils. Int J Syst Evol Microbiol. 2022;72:005546. https://doi.org/10.1099/ijsem.0.005546.
25. Liu G-H, Yang S, Tang R, Xie C-J, Zhou S-G. Genome Analysis and Description of Three Novel Diazotrophs *Geomonas* Species Isolated From Paddy Soils. Front Microbiol. 2021;12:801462. https://doi.org/10.3389/fmicb.2021.801462.
26. Masuda Y, Yamanaka H, Xu Z-X, Shiratori Y, Aono T, Amachi S, et al. Diazotrophic *Anaeromyxobacter* Isolates from Soils. Appl Environ Microbiol. 2020;86:e00956–20. https://doi.org/10.1128/AEM.00956-20.
27. Xu Z, Masuda Y, Hayakawa C, Ushijima N, Kawano K, Shiratori Y, et al. Description of Three Novel Members in the Family *Geobacteraceae*, *Oryzomonas japonicum* gen. nov., sp. nov., *Oryzomonas sagensis* sp. nov., and *Oryzomonas ruber* sp. nov. Microorganisms. 2020;8:634. https://doi.org/10.3390/microorganisms8050634.
28. Xu Z, Masuda Y, Itoh H, Ushijima N, Shiratori Y, Senoo K. *Geomonas oryzae* gen. nov., sp. nov., *Geomonas edaphica* sp. nov., *Geomonas ferrireducens* sp. nov., *Geomonas terrae* sp. nov., Four Ferric-Reducing Bacteria Isolated From Paddy Soil, and Reclassification of Three Species of the Genus *Geobacter* as Members of the Genus *Geomonas* gen. nov. Front Microbiol. 2019;10:2201. https://doi.org/10.3389/fmicb.2019.02201.
29. Xu Z, Masuda Y, Wang X, Ushijima N, Shiratori Y, Senoo K, et al. Genome-Based Taxonomic Rearrangement of the Order *Geobacterales* Including the Description of *Geomonas azotofigens* sp. nov. and *Geomonas diazotrophica* sp. nov. Front Microbiol. 2021;12:2715. https://doi.org/10.3389/fmicb.2021.737531.
30. Yang S, Liu G-H, Tang R, Han S, Xie C-J, Zhou S-G. Description of two nitrogen-fixing bacteria, *Geomonas fuzhouensis* sp. nov. and *Geomonas agri* sp. nov., isolated from paddy soils. Antonie Van Leeuwenhoek. 2022;115:435–44. https://doi.org/10.1007/s10482-021-01704-6.
31. Zhang Z, Xu Z, Masuda Y, Wang X, Ushijima N, Shiratori Y, et al. *Geomesophilobacter sediminis* gen. nov., sp. nov., *Geomonas propionica* sp. nov. and *Geomonas anaerohicana* sp. nov., three novel members in the family *Geobacterecae* isolated from river sediment and paddy soil. Syst Appl Microbiol. 2021;44:126233. https://doi.org/10.1016/j.syapm.2021.126233.
32. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, et al. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. Nat Microbiol. 2018;3:804–13. https://doi.org/10.1038/s41564-018-0176-9.
33. Jones CM, Graf DR, Bru D, Philippot L, Hallin S. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. The ISME Journal. 2013;7:417–26. https://doi.org/10.1038/ismej.2012.125.
34. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7:111–8. https://doi.org/10.1038/nmeth.1419.
35. Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R, et al. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput Biol Chem. 2008;32:452–7. https://doi.org/10.1016/j.compbiolchem.2008.07.021.
36. Mise K, Masuda Y, Senoo K, Itoh H. Undervalued Pseudo-*nifH* Sequences in Public Databases Distort Metagenomic Insights into Biological Nitrogen Fixers. mSphere. 2021;6:e00785–21. https://doi.org/10.1128/msphere.00785-21.
37. Strien J, Sanft J, Mall G. Enhancement of PCR Amplification of Moderate GC-Containing and Highly GC-Rich DNA Sequences. Mol Biotechnol. 2013;54:1048–54. https://doi.org/10.1007/s12033-013-9660-x.
38. Gaby JC, Buckley DH. A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase. PLoS One. 2012;9:e93883. https://doi.org/10.1371/journal.pone.0042149.
39. Kim M, Oh H-S, Park S-C, Chun J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014;64:346–51. https://doi.org/10.1099/ijs.0.059774-0.
40. Bazylinski DA, Dean AJ, Schuler D, Phillips EJP, Lovley DR. N_2-dependent growth and nitrogenase activity in the metal-metabolizing bacteria, *Geobacter* and *Magnetospirillum* species. Environ Microbiol. 2000;2:266–73. https://doi.org/10.1046/j.1462-2920.2000.00096.x.
41. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. Nucleic Acids Res. 2022;50:D387–90. https://doi.org/10.1093/nar/gkab1053.
42. Meyer F, Bagchi S, Chaterji S, Gerlach W, Grama A, Harrison T, et al. MG-RAST version 4—lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. Brief Bioinform. 2019;20:1151–9. https://doi.org/10.1093/bib/bbx105.
43. Mus F, Alleman AB, Pence N, Seefeldt LC, Peters JW. Exploring the alternatives of biological nitrogen fixation. Metallomics. 2018;10:523–38. https://doi.org/10.1039/C8MT00038G.
44. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50:D785–94. https://doi.org/10.1093/nar/gkab776.

45. Robson RL, Postgate JR. Oxygen and Hydrogen in Biological Nitrogen Fixation. Ann Rev Microbiol. 1980;34:183–207. https://doi.org/10.1146/annurev.mi.34.100180.001151.
46. Zhou J, Ma M, Guan D, Jiang X, Zhang N, Shu F, et al. Nitrogen has a greater influence than phosphorus on the diazotrophic community in two successive crop seasons in Northeast China. Sci Rep. 2021;11:6303. https://doi.org/10.1038/s41598-021-85829-8.
47. Wang H, He X, Zhang Z, Li M, Zhang Q, Zhu H, et al. Eight years of manure fertilization favor copiotrophic traits in paddy soil microbiomes. Eur J Soil Biol. 2021b;106:103352. https://doi.org/10.1016/j.ejsobi.2021.103352.
48. Drake JB, Weishampel JF. Multifractal analysis of canopy height measures in a longleaf pine savanna. For Ecol Manag. 2000;128:121–7. https://doi.org/10.1016/S0378-1127(99)00279-0.
49. Pi H-W, Lin J-J, Chen C-A, Wang P-H, Chiang Y-R, Huang C-C, et al. Origin and Evolution of Nitrogen Fixation in Prokaryotes. Molecular Biology and Evolution 39, msac181. 2022; https://doi.org/10.1093/molbev/msac181.
50. Raymond J, Siefert JL, Staples CR, Blankenship RE. The Natural History of Nitrogen Fixation. Mol Biol Evol. 2004;21:541–54. https://doi.org/10.1093/molbev/msh047.
51. Brown JR. Ancient horizontal gene transfer. Nat Rev Genet. 2003;4:121–32. https://doi.org/10.1038/nrg1000.
52. Wang H, Li X, Li X, Li F, Su Z, Zhang H. Community Composition and Co-Occurrence Patterns of Diazotrophs along a Soil Profile in Paddy Fields of Three Soil Types in China. Microb Ecol. 2021c;82:961–70. https://doi.org/10.1007/s00248-021-01716-9.
53. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, et al. Patterns and Processes of Microbial Community Assembly. Microbiol Mol Biol Rev. 2013;77:342–56. https://doi.org/10.1128/MMBR.00051-12.
54. Vellend M. Conceptual Synthesis in Community Ecology. Q Rev Biol. 2010;85:183–206. https://doi.org/10.1086/652373.
55. Fodelianakis S, Valenzuela-Cuevas A, Barozzi A, Daffonchio D. Direct quantification of ecological drift at the population level in synthetic bacterial communities. The ISME Journal. 2021;15:55–66. https://doi.org/10.1038/s41396-020-00754-4.
56. Xu X, Thornton PE, Post WM. A global analysis of soil microbial biomass carbon, nitrogen and phosphorus in terrestrial ecosystems. Glob Ecol Biogeogr. 2013;22:737–49. https://doi.org/10.1111/geb.12029.
57. Davidson NC, Fluet-Chouinard E, Finlayson CM. Global extent and distribution of wetlands: trends and issues. Mar Freshw Res. 2018;69:620. https://doi.org/10.1071/MF17019.
58. Wang X, Teng Y, Ren W, Li Y, Yang T, Chen Y, et al. Variations of Bacterial and Diazotrophic Community Assemblies throughout the Soil Profile in Distinct Paddy Soil Types and Their Contributions to Soil Functionality. mSystems. 2022;7:e01047–21. https://doi.org/10.1128/msystems.01047-21.
59. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. mSphere. 2021;6:e01202–20. https://doi.org/10.1128/mSphere.01202-20.
60. Kim DD, Park D, Yoon H, Yun T, Song MJ, Yoon S. Quantification of *nosZ* genes and transcripts in activated sludge microbiomes with novel group-specific qPCR methods validated with metagenomic analyses. Water Res. 2020;185:116261. https://doi.org/10.1016/j.watres.2020.116261.
61. Aird D, Ross MG, Chen W-S, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011;12:R18. https://doi.org/10.1186/gb-2011-12-2-r18.
62. Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, et al. Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. DNA Res. 2019;26:391–8. https://doi.org/10.1093/dnares/dsz017.
63. Sevim V, Lee J, Egan R, Clum A, Hundley H, Lee J, et al. Shotgun metagenome data of a defined mock community using Oxford Nanopore. PacBio and Illumina technologies Scientific Data. 2019;6:285. https://doi.org/10.1038/s41597-019-0287-z.
64. Soumare A, Diedhiou AG, Thuita M, Hafidi M, Ouhdouch Y, Gopalakrishnan S, et al. Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable Agriculture. Plants. 2020;9:1011. https://doi.org/10.3390/plants9081011.
65. Masuda Y, Shiratori Y, Ohba H, Ishida T, Takano R, Satoh S, et al. Enhancement of the nitrogen-fixing activity of paddy soils owing to iron application. Soil Sci Plant Nutr. 2021;67:243–7. https://doi.org/10.1080/00380768.2021.1888629.
66. Shen W, Long Y, Qiu Z, Gao N, Masuda Y, Itoh H, et al. Investigation of Rice Yields and Critical N Losses from Paddy Soil under Different N Fertilization Rates with Iron Application. Int J Environ Res Public Health. 2022;19:8707. https://doi.org/10.3390/ijerph19148707.
67. Choi J, Yang F, Stepanauskas R, Cardenas E, Garoutte A, Williams R, et al. Strategies to improve reference databases for soil microbiomes. The ISME Journal. 2017;11:829–34. https://doi.org/10.1038/ismej.2016.168.
68. Dash B, Nayak S, Pahari A, Nayak SK. Verrucomicrobia in Soil: An Agricultural Perspective. In: Frontiers in Soil and Environmental Microbiology. CRC Press; 2020. p. 37–46. https://doi.org/10.1201/9780429485794-4.
69. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Front Microbiol. 2016;7:744. https://doi.org/10.3389/fmicb.2016.00744.
70. Zhang Z, Masuda Y, Xu Z, Shiratori Y, Ohba H, Senoo K. Active nitrogen fixation by iron-reducing bacteria in rice paddy soil and its further enhancement by iron application. Appl Sci. 2023;13:8156. https://doi.org/10.3390/app13148156.
71. Coates JD, Phillips EJ, Lonergan DJ, Jenter H, Lovley DR. Isolation of *Geobacter* species from diverse sedimentary environments. Appl Environ Microbiol. 1996;62:1531–6. https://doi.org/10.1128/aem.62.5.1531-1536.1996.
72. Lovley DR, Giovannoni SJ, White DC, Champine JE, Phillips EJP, Gorby YA, et al. *Geobacter metallireducens* gen. nov. sp. nov., a microorganism capable of coupling the complete oxidation of organic compounds to the reduction of iron and other metals. Arch Microbiol. 1993;159:336–44. https://doi.org/10.1007/BF00290916.
73. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9. https://doi.org/10.1101/gr.074492.107.
74. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: Integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49:D545–51. https://doi.org/10.1093/nar/gkaa970.
75. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119. https://doi.org/10.1186/1471-2105-11-119.
76. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2. https://doi.org/10.1093/bioinformatics/btz859.
77. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436.
78. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One. 2010;5:e9490. https://doi.org/10.1371/journal.pone.0009490.
79. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36:1925–7. https://doi.org/10.1093/bioinformatics/btz848.
80. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
81. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. Nat Microbiol. 2016;1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.
82. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6. https://doi.org/10.1093/nar/gkab301.
83. Takada-Hoshino Y, Matsumoto N. An Improved DNA Extraction Method Using Skim Milk from Soils That Strongly Adsorb DNA. Microbes Environ. 2004;19:13–9. https://doi.org/10.1264/jsme2.19.13.
84. Arita M, Karsch-Mizrachi I, Cochrane G. The international nucleotide sequence database collaboration. Nucleic Acids Res. 2021;49:D121–4. https://doi.org/10.1093/nar/gkaa967.
85. Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M, et al. The metagenomics RAST server – a public resource for the automatic

phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9:386. https://doi.org/10.1186/1471-2105-9-386.

86. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat Microbiol. 2018;3:470–80. https://doi.org/10.1038/s41564-018-0129-3.
87. Angle JC, Morin TH, Solden LM, Narrowe AB, Smith GJ, Borton MA, et al. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. Nat Commun. 2017;8:1567. https://doi.org/10.1038/s41467-017-01753-4.
88. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, et al. Structure and function of the global topsoil microbiome. Nature. 2018;560:233–7. https://doi.org/10.1038/s41586-018-0386-6.
89. Berkelmann D, Schneider D, Meryandini A, Daniel R. Unravelling the effects of tropical land use conversion on the soil microbiome. Environmental Microbiome. 2020;15:5. https://doi.org/10.1186/s40793-020-0353-3.
90. Black EM, Just CL. The Genomic Potentials of NOB and Comammox Nitrospira in River Sediment Are Impacted by Native Freshwater Mussels. Front Microbiol. 2018;9:2061. https://doi.org/10.3389/fmicb.2018.02061.
91. Cania B, Vestergaard G, Krauss M, Fliessbach A, Schloter M, Schulz S. A long-term field experiment demonstrates the influence of tillage on the bacterial potential to produce soil structure-stabilizing agents such as exopolysaccharides and lipopolysaccharides. Environmental Microbiome. 2019;14:1. https://doi.org/10.1186/s40793-019-0341-7.
92. Cha G, Meinhardt KA, Orellana LH, Hatt JK, Pannu MW, Stahl DA, et al. The influence of alfalfa-switchgrass intercropping on microbial community structure and function. Environ Microbiol. 2021;23:6828–43. https://doi.org/10.1111/1462-2920.15785.
93. Chen Y-P, Liaw L-L, Kuo J-T, Wu H-T, Wang G-H, Chen X-Q, et al. Evaluation of synthetic gene encoding α-galactosidase through metagenomic sequencing of paddy soil. J Biosci Bioeng. 2019;128:274–82. https://doi.org/10.1016/j.jbiosc.2019.03.006.
94. Chu BTT, Petrovich ML, Chaudhary A, Wright D, Murphy B, Wells G, et al. Metagenomics Reveals the Impact of Wastewater Treatment Plants on the Dispersal of Microorganisms and Genes in Aquatic Sediments. Appl Environ Microbiol. 2018;84 https://doi.org/10.1128/AEM.02168-17.
95. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature. 2018;558:440–4. https://doi.org/10.1038/s41586-018-0207-y.
96. Hartman WH, Ye R, Horwath WR, Tringe SG. A genomic perspective on stoichiometric regulation of soil carbon cycling. The ISME Journal. 2017;11:2652–65. https://doi.org/10.1038/ismej.2017.115.
97. Huber DH, Ugwuanyi IR, Malkaram SA, Montenegro-Garcia NA, Lhilhi Noundou V, Chavarria-Palma JE. Metagenome Sequences of Sediment from a Recovering Industrialized Appalachian River in West Virginia. Genome Announcements. 2018;6:e00350–18. https://doi.org/10.1128/genomeA.00350-18.
98. Jiang H, Zhou R, Zhang M, Cheng Z, Li J, Zhang G, et al. Exploring the differences of antibiotic resistance genes profiles between river surface water and sediments using metagenomic approach. Ecotoxicol Environ Saf. 2018;161:64–9. https://doi.org/10.1016/j.ecoenv.2018.05.044.
99. Johnston ER, Rodriguez-R LM, Luo C, Yuan MM, Wu L, He Z, et al. Metagenomics Reveals Pervasive Bacterial Populations and Reduced Community Diversity across the Alaska Tundra Ecosystem. Front Microbiol. 2016;7:579. https://doi.org/10.3389/fmicb.2016.00579.
100. Li H-Y, Wang H, Wang H-T, Xin P-Y, Xu X-H, Ma Y, et al. The chemodiversity of paddy soil dissolved organic matter correlates with microbial community at continental scales. Microbiome. 2018;6:187. https://doi.org/10.1186/s40168-018-0561-x.
101. Li Y, Tremblay J, Bainard LD, Cade-Menun B, Hamel C. Long-term effects of nitrogen and phosphorus fertilization on soil microbial community structure and function under continuous wheat production. Environ Microbiol. 2020;22:1066–88. https://doi.org/10.1111/1462-2920.14824.
102. Links MG, Dumonceaux TJ, McCarthy EL, Hemmingsen SM, Topp E, Town JR. CaptureSeq: Hybridization-Based Enrichment of *cpn60* Gene Fragments Reveals the Community Structures of Synthetic and Natural Microbial Ecosystems. Microorganisms. 2021;9:816. https://doi.org/10.3390/microorganisms9040816.
103. Liu Y-R, Johs A, Bi L, Lu X, Hu H-W, Sun D, et al. Unraveling Microbial Communities Associated with Methylmercury Production in Paddy Soils. Environ Sci Technol. 2018;52:13110–8. https://doi.org/10.1021/acs.est.8b03052.
104. Ma B, Zhao K, Lv X, Su W, Dai Z, Gilbert JA, et al. Genetic correlation network prediction of forest soil microbial functional organization. The ISME Journal. 2018;12:2492–505. https://doi.org/10.1038/s41396-018-0232-8.
105. Neal AL, Hughes D, Clark IM, Jansson JK, Hirsch PR. Microbiome Aggregated Traits and Assembly Are More Sensitive to Soil Management than Diversity. mSystems. 2021;6 https://doi.org/10.1128/mSystems.01056-20.
106. Nelkner J, Henke C, Lin TW, Pätzold W, Hassa J, Jaenicke S, et al. Effect of Long-Term Farming Practices on Agricultural Soil Microbiome Members Represented by Metagenomically Assembled Genomes (MAGs) and Their Predicted Plant-Beneficial Genes. Genes. 2019;10:424. https://doi.org/10.3390/genes10060424.
107. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. Year-Round Shotgun Metagenomes Reveal Stable Microbial Communities in Agricultural Soils and Novel Ammonia Oxidizers Responding to Fertilization. Appl Environ Microbiol. 2018;84 https://doi.org/10.1128/AEM.01646-17.
108. Ouyang Y, Norton JM. Short-Term Nitrogen Fertilization Affects Microbial Community Composition and Nitrogen Mineralization Functions in an Agricultural Soil. Appl Environ Microbiol. 2020;86:516–8. https://doi.org/10.1128/AEM.02278-19.
109. Paungfoo-Lonhienne C, Wang W, Yeoh YK, Halpin N. Legume crop rotation suppressed nitrifying microbial community in a sugarcane cropping soil. Sci Rep. 2017;7:16707. https://doi.org/10.1038/s41598-017-17080-z.
110. Romanowicz KJ, Crump BC, Kling GW. Rainfall Alters Permafrost Soil Redox Conditions, but Meta-Omics Show Divergent Microbial Community Responses by Tundra Type in the Arctic. Soil Systems. 2021;5:17. https://doi.org/10.3390/soilsystems5010017.
111. Sukhum KV, Vargas RC, Boolchandani M, D'Souza AW, Patel S, Kesaraju A, et al. Manure Microbial Communities and Resistance Profiles Reconfigure after Transition to Manure Pits and Differ from Those in Fertilized Field Soil. mBio. 2021;12 https://doi.org/10.1128/mBio.00798-21.
112. Suttner B, Johnston ER, Orellana LH, Rodriguez-R LM, Hatt JK, Carychao D, et al. Metagenomics as a Public Health Risk Assessment Tool in a Study of Natural Creek Sediments Influenced by Agricultural and Livestock Runoff: Potential and Limitations. Appl Environ Microbiol. 2020;86 https://doi.org/10.1128/AEM.02525-19.
113. Wang J, Long Z, Min W, Hou Z. Metagenomic analysis reveals the effects of cotton straw–derived biochar on soil nitrogen transformation in drip-irrigated cotton field. Environ Sci Pollut Res. 2020;27:43929–41. https://doi.org/10.1007/s11356-020-10267-4.
114. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, et al. Genome-centric view of carbon processing in thawing permafrost. Nature. 2018;560:49–54. https://doi.org/10.1038/s41586-018-0338-1.
115. Wu D, Zhao Y, Cheng L, Zhou Z, Wu Q, Wang Q, et al. Activity and structure of methanogenic microbial communities in sediments of cascade hydropower reservoirs, Southwest China. Sci Total Environ. 2021;786:147515. https://doi.org/10.1016/j.scitotenv.2021.147515.
116. Xiao K-Q, Li B, Ma L, Bao P, Zhou X, Zhang T, et al. Metagenomic profiles of antibiotic resistance genes in paddy soils from South China. FEMS Microbiol Ecol. 2016;92:fiw023. https://doi.org/10.1093/femsec/fiw023.
117. Xue Y, Jonassen I, Øvreås L, Taş N. Bacterial and Archaeal Metagenome-Assembled Genome Sequences from Svalbard Permafrost. Microbiology Resource Announcements. 2019;8 https://doi.org/10.1128/MRA.00516-19.
118. Yu J, Deem LM, Crow SE, Deenik J, Penton CR. Comparative Metagenomics Reveals Enhanced Nutrient Cycling Potential after 2 Years of Biochar Amendment in a Tropical Oxisol. Appl Environ Microbiol. 2019;85 https://doi.org/10.1128/AEM.02957-18.
119. Yurgel SN, Nearing JT, Douglas GM, Langille MGI. Metagenomic Functional Shifts to Plant Induced Environmental Changes. Front Microbiol. 2019;10:1682. https://doi.org/10.3389/fmicb.2019.01682.

120. Zhang C, Song Z, Zhuang D, Wang J, Xie S, Liu G. Urea fertilization decreases soil bacterial diversity, but improves microbial biomass, respiration, and N-cycling potential in a semiarid grassland. Biol Fertil Soils. 2019;55:229–42. https://doi.org/10.1007/s00374-019-01344-z.
121. Zheng Z, Li L, Makhalanyane TP, Xu C, Li K, Xue K, et al. The composition of antibiotic resistance genes is not affected by grazing but is determined by microorganisms in grassland soils. Sci Total Environ. 2021;761:143205. https://doi.org/10.1016/j.scitotenv.2020.143205.
122. Courtot M, Gupta D, Liyanage I, Xu F, Burdett T. BioSamples database: FAIRer samples metadata to accelerate research data management. Nucleic Acids Res. 2022;50:D1500–7. https://doi.org/10.1093/nar/gkab1046.
123. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1. https://doi.org/10.1093/bioinformatics/btq461.
124. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. https://doi.org/10.1038/nmeth.3869.
125. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One. 2016;11:e0163962. https://doi.org/10.1371/journal.pone.0163962.
126. R Core Team, 2021. R: A Language and Environment for Statistical Computing.
127. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package; 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.