Phylogenetics and environmental distribution of nitric oxide-forming nitrite reductases reveal their distinct functional and ecological roles

Grace Pold[1], Germán Bonilla-Rosso[1,‡], Aurélien Saghaï[1], Marc Strous[2], Christopher M. Jones[1], Sara Hallin[1,*]

[1]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden
[2]Department of Earth, Energy, and Environment, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

*Corresponding author: Sara Hallin, Swedish University of Agricultural Sciences, Uppsala Biocenter, Department of Forest Mycology and Plant Pathology, Box 7026, 750 07 Uppsala, Sweden. Email: sara.hallin@slu.se
‡Present address: Method Development and Analytics—Molecular Ecology, Agroscope, Zurich, Switzerland.

Abstract

The two evolutionarily unrelated nitric oxide-producing nitrite reductases, NirK and NirS, are best known for their redundant role in denitrification. They are also often found in organisms that do not perform denitrification. To assess the functional roles of the two enzymes and to address the sequence and structural variation within each, we reconstructed robust phylogenies of both proteins with sequences recovered from 6973 isolate and metagenome-assembled genomes and identified 32 well-supported clades of structurally distinct protein lineages. We then inferred the potential niche of each clade by considering other functional genes of the organisms carrying them as well as the relative abundances of each *nir* gene in 4082 environmental metagenomes across diverse aquatic, terrestrial, host-associated, and engineered biomes. We demonstrate that Nir phylogenies recapitulate ecology distinctly from the corresponding organismal phylogeny. While some clades of the nitrite reductase were equally prevalent across biomes, others had more restricted ranges. Nitrifiers make up a sizeable proportion of the nitrite-reducing community, especially for NirK in marine waters and dry soils. Furthermore, the two reductases showed distinct associations with genes involved in oxidizing and reducing other compounds, indicating that the NirS and NirK activities may be linked to different elemental cycles. Accordingly, the relative abundance and diversity of NirS versus NirK vary between biomes. Our results show the divergent ecological roles NirK and NirS-encoding organisms may play in the environment and provide a phylogenetic framework to distinguish the traits associated with organisms encoding the different lineages of nitrite reductases.

Keywords: denitrification, nitrite reductase, metagenomes, comparative genomics, phylogenetics

Introduction

Nitrite is a key intermediate in the global nitrogen (N) cycle, as it is either produced or consumed in most inorganic N transformations. In denitrification, the major route of N loss from the biosphere to the atmosphere, nitrite is reduced to gaseous nitric oxide (NO) by one of two evolutionarily distinct nitrite reductases (hereafter, referred as Nir): the heme-coordinating cytochrome cd_1 NirS and the multicopper-oxidase NirK. In addition to structural differences, NirS requires more helper proteins to assemble [1], and it is more likely to co-occur with other genes in the denitrification pathway [2]. Nonetheless, the nitrite-reducing capabilities of NirS and NirK are biochemically redundant in denitrifying organisms, and genes for the two enzymes are rarely found in the same genome [2], although each can compensate for the absence of the other [3, 4]. However, the ratio of genes encoding NirK versus NirS within whole communities varies both between habitats and in response to environmental change [5–13], suggesting divergent ecological roles of the organisms encoding them. This may be in part associated with the diversity of other ecological processes nitrite reductases can be associated with in addition to denitrification, including incomplete denitrification (nitrate/nitric oxide reduction or nitrous oxide reduction [2]); ammonia/hydroxylamine [14, 15], sulphur [16], iron [17], and manganese oxidation [18]; nitric oxide pool replenishment for hydrazine synthase in anaerobic ammonium oxidation (anammox) [19, 20]; nitrite [21, 22] and selenite detoxification [23]; pathogenesis [24]; and magnetite [25, 26] and oxygen generation [27]. These alternative roles may explain the differences in the NirK and NirS communities across biomes [28, 29]. While earlier work suggests a complex evolutionary history for NirS and NirK [30], the massive increase of sequencing data over the past decade offers new possibilities to gain insights into the structure–function relationships of these proteins within an eco-evolutionary context.

Here, we leveraged publicly available genomes and environmental metagenomes to examine the overall diversity of both NirS and NirK to evaluate the hypothesis that the distribution of Nir types, clades, and subclades across biomes can be explained by environmental conditions (Fig. 1). Specifically, we posit that NirS dominates under conditions more consistently suitable for denitrification, as suggested by both its greater enzyme production

cost compared to NirK [10] and its prevalence within organisms capable of completely reducing nitrite to dinitrogen ("complete denitrifiers"; [2]). We further hypothesize that the sequence variation in Nir proteins is linked to ecological differences rather than solely shared organism ancestry [31]. This is primarily rooted in the observation that NirS and particularly NirK phylogenies are incongruent with the 16S rRNA phylogeny [2, 30, 32] and do not reflect the evolutionary history of the organisms harbouring these genes at a finer taxonomic level. To test these hypotheses, we applied a phylogenetic approach to analyse the NirK and NirS sequences obtained by screening >1 000 000 assemblies of isolate and metagenome-assembled genomes, and then used structural features to further categorize NirK and NirS sequence variants into clades. We subsequently identified associations of each Nir clade with the presence of other functional genes found in the genomes of organisms within different clades to infer other terminal electron acceptors and donors that organisms may utilize (i.e. redox traits). This was complemented by a global analysis in which we classified environmental NirS and NirK fragments from 4082 publicly available metagenomes spanning major terrestrial, aquatic, engineered, and host-associated biomes (Table S1) and determined the environmental associations differentiating NirK and NirS as well as the clades comprising each gene. This study provides an unprecedented survey of the diversity and environmental distribution of NO-forming nitrite reductases and shows that the Nir phylogenies recapitulate ecology beyond organismal phylogeny. Such an enhanced understanding of the ecology of nitrite reducers will enable an improved interpretation of future metagenomes regarding this essential function.

Materials and methods
Phylogenetic reconstruction of nitric oxide-forming nitrite reductase homologues in genomes

Phylogenies were constructed using a combination of hidden Markov model (HMM)-based approaches and manual curation (Fig. 1A). All bacteria and archaea (7 October 2021), fungi (21 November 2021), and plant and protozoa genome assemblies (19 January 2022) were downloaded from NCBI and were searched for NirK and NirS using "hmmer" [33] against an HMM built from the alignments of the respective protein from Graf *et al.* [2]. Eighteen NirK sequences inferred from foraminifera transcriptomes were also included [34]. After translating the DNA sequences into amino acids and removing identical sequences using CD-HIT (v 4.6; [35, 36]), initial maximum likelihood (ML) trees were reconstructed using FastTree v. 2.1.11 [37].

The resulting Nir trees were then manually pruned to reduce the number of tips while maintaining the breadth of sequence divergence. Sequences were removed if they came from genome assemblies marked as "contaminated" in GenBank, or if they had >5% contamination and/or <90% (80% for eukaryotes) completeness according to BUSCO (v. 5.3.1; [38]. Some lower quality assemblies were retained in poorly sampled regions of the phylogeny. We also excluded NirK sequences identified as occurring on prokaryote-like contigs in eukaryotic assemblies by mmseqs2-taxonomy (v.14; [39]) against the UniRef50 database (release 2021_04; [40]). Taxonomy was assigned to prokaryotic assemblies using GTDB-TK (v. 1.5.0, GTDB R202; [41]), while taxonomy reported with the assembly in NCBI was assigned to eukaryotes. Alignments underwent a final round of manual editing and alignment following the structural information available for NirS [42, 43] and NirK [44] using ARB v. 7.0 [45]. Positions with <5% similarity were excluded from the alignment

prior to phylogeny construction, leaving alignments of 573 and 469 amino acids for NirK and NirS, respectively. The best evolutionary model for each protein was determined using modelfinder (LG + F + R10 for NirK and LG + F + R9 for NirS), and phylogenies were subsequently built using IQ-TREE v. 2.1.3 [46]. Node support was determined using the ultrafast bootstrap (N = 1000; [46]) and the Shimodaira–Hasegawa approximate likelihood ratio test [47]. Clades were defined with node support values SH-aLRT $\geq$80% and UFboot $\geq$95%. In addition to these statistical support criteria, we used protein structure, gene neighbourhood, and previous knowledge on the ecology and/or taxonomy of the encoding organisms to define clades. The clades were named following existing literature where possible [28, 29], reserving Clade 1 of each protein for the "canonical clade" and letters for sub clades within each larger clade. Trees and associated metadata were plotted using iTOL v5 [48].
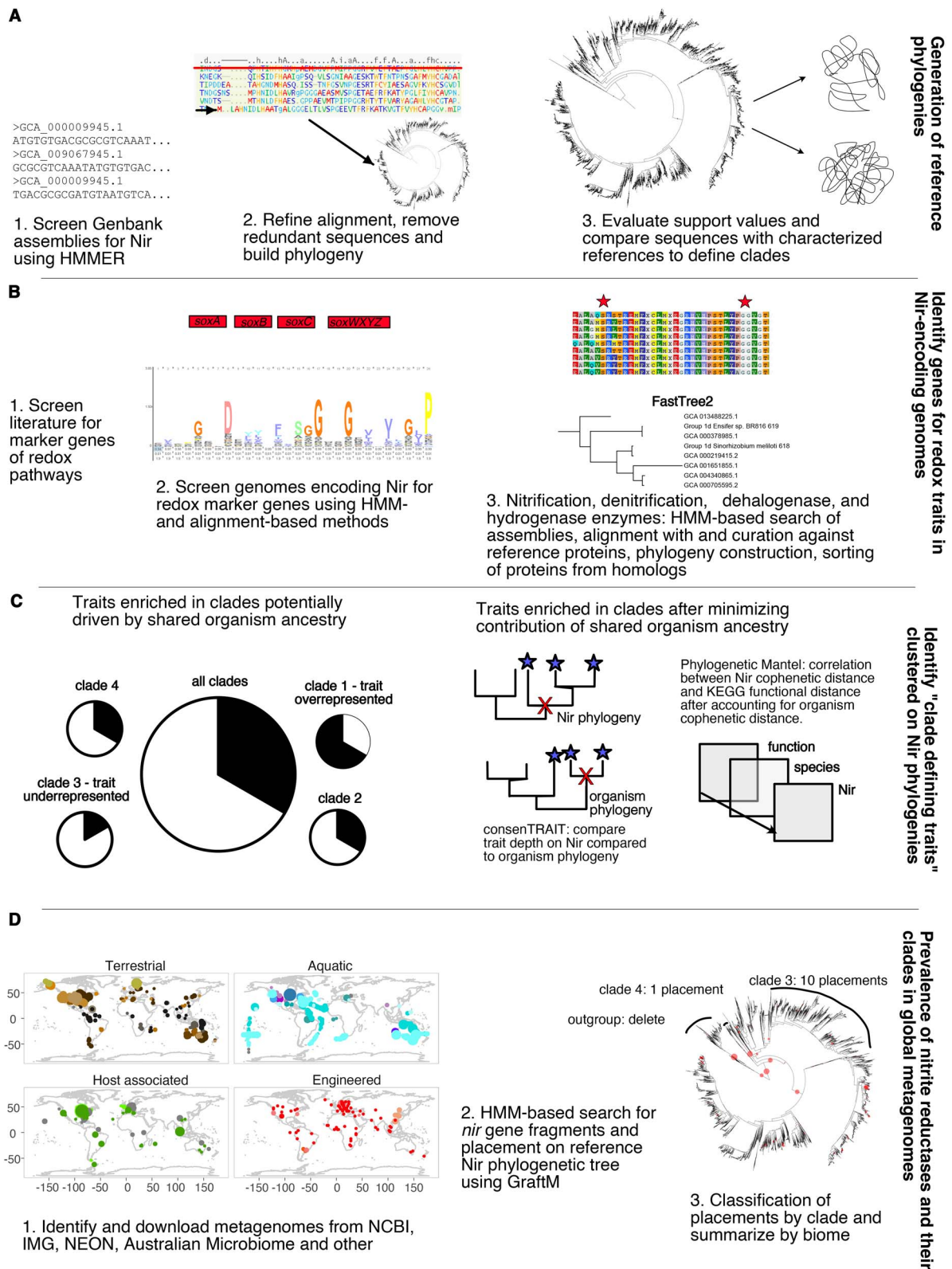
Phylogenetic reconstruction of organism history

We built organism phylogenies for bacteria and archaea using a conserved set of 16 ribosomal proteins encoded by single copy genes [49], as facilitated by GToTree v1.7.07 [50]. We excluded eukaryotic sequences from these organism phylogenies and traits analysis due to the lack of reliable protein predictions. Genomes in which fewer than 8 of the 16 targeted ribosomal protein genes were identified were also excluded from the alignment. We subsequently used modelfinder to infer the best amino acid matrix for each partition (Table S2) and identified the most likely tree using IQ-TREE. Due to computing constraints, we manually generated 100 bootstraps of the starting alignment by sampling within each partition and generated an approximate ML tree of each alignment using the WG model in FastTree. Zero-length branches were set to 10× smaller than the smallest edge in the tree to allow consenTRAIT to run. The final phylogeny included 5518 tips for NirK and 528 for NirS.

Evaluating Nir-niche and Nir-organism phylogenetic congruence

We used a combination of broad metabolic and gene-specific trait approaches to test for ecologically defined clades in the Nir trees. For the broad metabolic characterization, we used EnrichM v.0.6.3 [51] against the KEGG v.10 database to identify orthologues and KEGG pathways that are >90% complete in the assemblies comprising each tree [52] (Fig. 1B). The resulting gene or pathway presence–absence matrix was used to generate Sorensen dissimilarity matrices and was subjected to phylogenetic Mantel tests [53] to evaluate whether the dissimilarity in Nir sequence (i.e. cophenetic distance) correlated with the overall dissimilarity in KEGG orthologue composition while accounting for common organismic history (Fig. 1C). Here, the organism tree was used to constrain the permutations so that more closely related taxa were more likely to switch traits than more distantly related taxa. We used Spearman's rho as our metric of correlation to account for possible monotonic, nonlinear associations.

Our gene-specific approach focused on the potential to reduce alternative electron acceptors and oxidize inorganic electron donors, reflecting the range of redox states that organisms experience in the environment. We used a combination of HMM searches, phylogenetic reconstruction, and manual inspection of alignments to identify the genes involved in the reduction and oxidation of inorganic compounds (Supplemental Methods). To detect the physiological traits associated with Nir-based nitrite reduction, we searched for traits that were more conserved along the Nir phylogeny than along the organism phylogeny (Fig. 1C). We determined the extent of phylogenetic conservation of the

Figure 1. Methods overview diagram. A) generation of reference alignment B) phylogenetic reconstruction C) evaluation of phylogenetic congruence between Nir and organism phylogenies D) screening metagenomes for *nir* gene fragments. HMM logo in panel b was generated in Skylign (http://skylign.org/).

traits using consenTRAIT's τ_D statistic, which is the phylogenetic depth at which 90% of tips share a trait [54]. We calculated an effect size based on Glass' D [55] for each trait by comparing the observed τ_D across 100 bootstrap trees to the τ_D expected under the null model where traits are distributed at random on the phylogeny. This allowed us to account for both the uncertainty in the phylogeny structure and undersampled tips as well as compare across phylogenies with different scales. Traits with

larger positive effect sizes on the Nir than organism phylogeny were identified as specifically associated with Nir evolution. All redox traits were subsequently subjected to chi-square and *post hoc* chi-square tests to identify those clades that were more commonly associated with a given trait than expected by chance ($\alpha = 0.05$). These were reported as "clade defining traits" (Fig. 1C). All analyses were completed in R v.4.2.0 [56]

Metagenomic analyses

We searched for publicly available metagenomes sequenced using Illumina platforms that were $\geq 100\,000$ reads, ≥ 150 nucleotides long, and accompanied by sample metadata, focusing on large sequencing projects that had used a standardized metadata collection protocol. The resulting 4082 metagenomes were categorized into four main groups: terrestrial, aquatic (both water and sediments), host-associated (plant and animal, excluding humans), and engineered (Table S1 and Fig. 1D). Terrestrial metagenomes were then assigned to biomes following The Nature Conservancy Terrestrial Ecoregions [57] using the R packages sp v.1.4-5 [58], rgdal v.1.5-23 [59], and rgeos v.0.5-5 [60]. Agricultural soils were excluded from the biome-based approach and were instead categorized as croplands. Marine metagenomes were assigned to biomes based on latitude (polar: latitude >60°; mid-latitude: 30–60°; low-latitude: 0–30°) and freshwater environments were split based on author descriptions of sites. Quantitative metadata were converted into the same units where possible to enable comparison across studies.

Phylogenetic placement of environmental *nir* reads

To identify the *nirK* and *nirS* fragments in metagenomes, we used GraftM v.0.13.1 [61] which uses an HMM search of translated amino acid sequences followed by placement on a phylogenetic tree to assign reads to clades or taxa (Fig. 1D). A package was built for each protein after calculating the substitution model parameters in "RAxML" v.7.7.2 [62] and rerooting the respective phylogenies in FigTree v.1.4.4 [63]. Comparable outputs from GraftM require that read length is consistent among samples and therefore only the first 150 nucleotides were used. We tested that these parameters were appropriate by completing GraftM on >200 000 150 nt fragments of full-length *nirK* and *nirS* genes and their homologues curated from a large MAG database [64]. Sensitivity was calculated as the fraction of ingroup reads detected by the process, and it was 76% for *nirK* and 93% for *nirS*. Specificity was defined as the fraction of outgroup reads correctly placed in the outgroup, and it was 97% for *nirK* and 100% for *nirS*.

We then used the gappa v.0.8.0 [65] and pplacer v.1.1.19 [66] phylogenetic placement software suites to retain reads where at least 95% of the placement likelihood fell within ingroups (for overall counts) or specific clades (for clade counts) and a likelihood of zero in the outgroup. Placements in the halophilic archaea NirS-like clade, which lacks a cytochrome c domain, were excluded from total *nirS* counts. We then evaluated the relative dominance of *nirK* versus *nirS* in biomes using a method which accounts for both the difference in the relative abundance of these genes and the overall prevalence of *nir* within metagenomes, including those for which *nirK* and/or *nirS* were not detected.

$$\left(\delta nir = 10^9 * \left(\frac{nirK - nirS}{\text{total reads} * 150\text{bp}} \right) \right)$$

To evaluate how the Nir diversity varies across habitats, we calculated the balance-weighted phylogenetic diversity (BWPD)

using guppy with placements weighted according to their relative abundance [66, 67]. We performed detrended correspondence analysis using vegan (v. 2.6-4; [68]) to visualize biome-level differences in the clade composition for each Nir type using the average of 100 rarefactions at a depth of 15 placements. We also calculated phylogenetic edge correlations [65] for a finer-scale examination of how nitrite reductase carrying communities differ over environmental gradients known to affect nitrite reductase-carrying communities. For soils, we examined the ammonium, nitrate, soil organic carbon (SOC), soil moisture, copper, and iron content, and pH and soil texture (percent sand). For marine water column samples, we evaluated correlations with ammonium, nitrate and nitrite concentrations, dissolved oxygen, chlorophyll, temperature, and sampling depth. Except for δnir, these analyses used metagenomes with >15 placements and for which we had quantitative metadata (*nirS*: $n = 224$ aquatic, $n = 1089$ soils; *nirK*: $n = 588$ aquatic, $n = 1696$ soils) because species dissimilarity methods did not converge with fewer placements. Abundance and alpha diversity of *nirK* and *nirS* were compared between biomes using Benjamini–Hochberg FDR-corrected pairwise comparison of ranks following Kruskall–Wallis.

Results

### Structure of Nir phylogenies

nirS was detected in the genomes of 24 bacterial phyla (predominantly Proteobacteria) and Hydrothermarchaeota, with the final phylogeny consisting of 540 NirS sequences (Fig. 2). The two haem d_1 biosynthesis proteins, NirN and NirF, formed the external and internal outgroups, respectively, to NirS which could be distinguished based on previously described motifs [69] (Fig. S1 and Table S3). A Halobacteriota cytochrome d protein lacking the c-type cytochrome domain of NirS formed an additional outgroup (Supplementary Text). Ingroup sequences fell into 15 clades and could be partially distinguished by the presence of the *nirBEFJNT* genes encoding NirS heme biosynthesis and assembly proteins (Fig. 2 and Table S3). Sixteen percent of the assemblies represented by sequences in the NirS phylogeny encoded multiple copies of NirS, with 85% of these assemblies encoding NirS derived from distinct clades (Table S4).

The NirK phylogeny (Fig. 3) contained much more structurally and taxonomically diverse sequences than the NirS phylogeny, with 6422 NirK sequences derived from all three domains of life, including protists and fungi within eukaryotes, Halobacteriota and Thermoproteota within archaea, and 30 bacterial phyla. The outgroup consisted of 367 multicopper oxidase sequences derived from Cyanobacteria and Thermoproteota and could be distinguished from NirK based on previously identified motifs within the catalytic sites (Fig. S2 and Table S5; [44]). Ingroup sequences fell into 17 clades encompassing 75% of the sequences, with 90% of the remaining sequences occurring in the large, poorly supported region of the phylogeny, including AniA-like NirK sequences. Ten percent of the assemblies represented by sequences in the NirK phylogeny encoded multiple copies of NirK, with 44% of these derived from different clades (Table S6). Co-occurrence of both *nirK* and *nirS* genes was observed in 5.3% of all genome assemblies represented in either tree (Tables S4 and S6).

Environmental distribution and functional potential of NirS clades

The functional potential and environmental distribution of *nirS* sequences were nonrandomly distributed on its phylogeny. Overall functional dissimilarity based on KEGG orthologues and
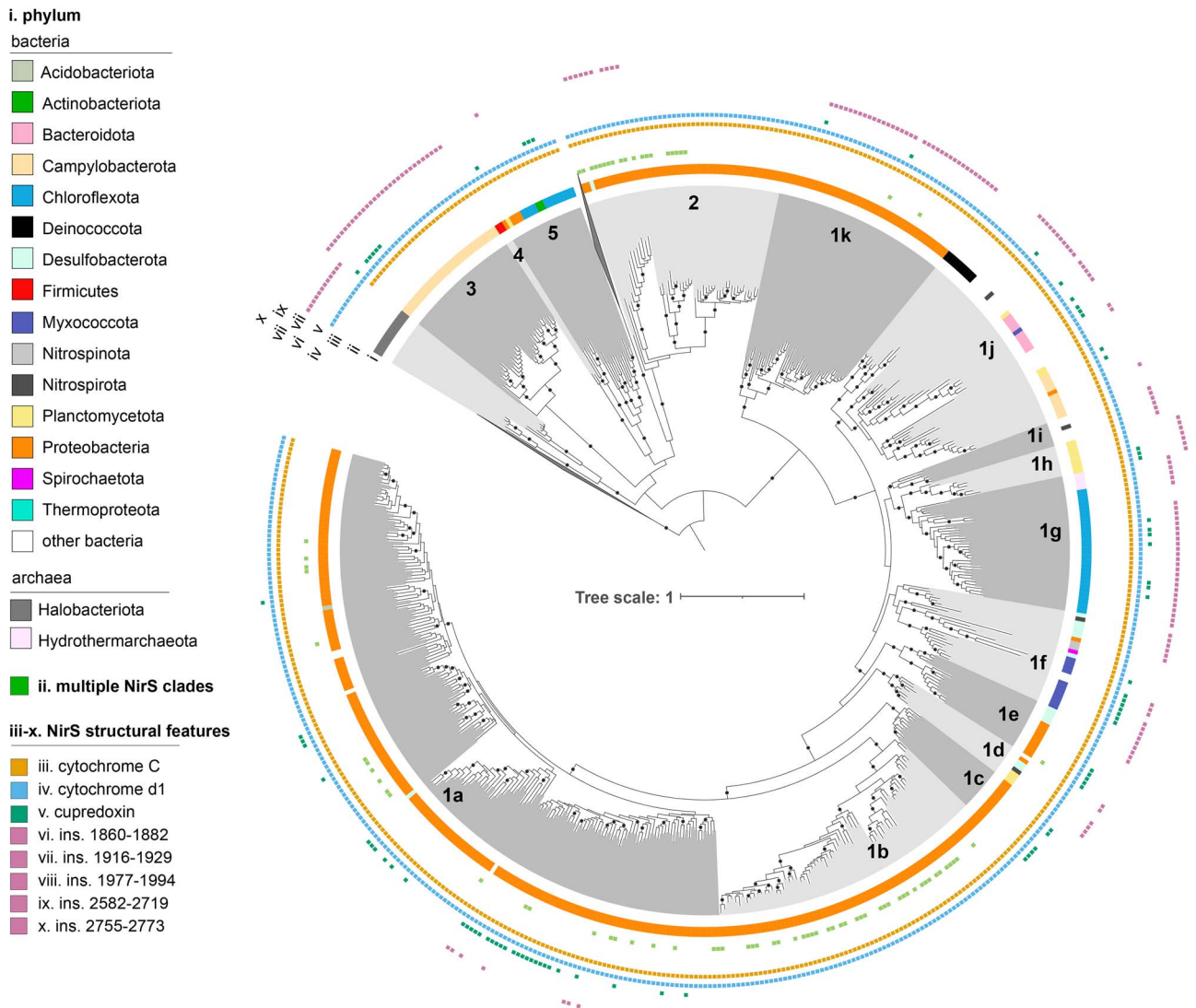
Figure 2. Maximum-likelihood NirS phylogeny. The phylogeny was inferred from 542 full-length NirS amino acid sequences using the LG+F+R9 model in IQ-TREE. Black circles on tree indicate nodes with SH-aLRT ≥ 80% and Ufboot ≥ 95% support. Tips of tree are shaded according to clade, and data rings denote (i) phylum determined using GTDB-TK; (ii) presence of secondary *nirS* copies from other clades; (iii–x) select NirS structural features. The two outgroups, consisting of NirN and NirF sequences, are collapsed for clarity. Clades are numbered and lettered following Wei *et al.* [29] and Bonilla-Rosso *et al.* [28] to the degree possible, starting with the "canonical" Clade 1a [48].

KEGG pathways correlated with the NirS sequence dissimilarity (phylogenetic Mantel test $r = 0.21$ and $r = 0.26$, respectively; $P < 0.001$). The genomic potential to oxidize or reduce specific compounds ("redox traits") was likewise heterogeneously distributed on the NirS phylogeny, with hydrogen and iron oxidation, complete denitrification, and the quinol-dependent nitric oxide reduction being more clustered than expected, given organism ancestry (consenTRAIT analysis; Table S7). Biome-level differences in the NirS community composition were similarly apparent (Fig. 4A).Specific NirS clades thus appeared to be functionally distinct and were often associated with specific habitats. The well-characterized proteobacterial Clade 1a made up the majority of reads in many biomes, particularly in engineered biomes but also in croplands and the endosphere and rhizosphere (Fig. S3 and Table S8). This clade shows an overall greater relative abundance under moist, organic matter- or ammonia-rich and low sand content soils, and in marine waters that were warmer, had lower dissolved oxygen, or had higher ammonium and chlorophyll content though not without

exception (Fig. 5A and B). Its sister Clade 1b, also predominantly consisting of proteobacterial sequences, showed much weaker correlations with the environmental variables assessed, largely because it was not prevalent in soils or ocean waters (Fig. S4). Rather, this clade was most prevalent in geysers (median 44%) and engineered biomes (Fig. S3) and was characterized by a greater-than-expected association with the potential for iron oxidation and complete denitrification (Fig. 6B and Table S9). Clades 1c (Thiobacillaceae) and 3 (Campylobacterales) consisted of sequences derived from sulphur- and hydrogen-oxidizing lithotrophs, but with different biome preferences; the former was abundant in geysers (24% median relative abundance), while the latter was most abundant in ponds and lakes (15% median relative abundance). Clades derived from sulphur-oxidizing lithotrophs were dominated by marine proteobacterial Clade 1k in the sea floor (8%) and marshes (6%). Clade 1f consisted of sequences derived from Myxococcota among other phyla, and it comprised 3%–11% of *nirS* sequences in saltwater habitats (Fig. S3). Clade 1j consisted of sequences derived mostly from thermophilic
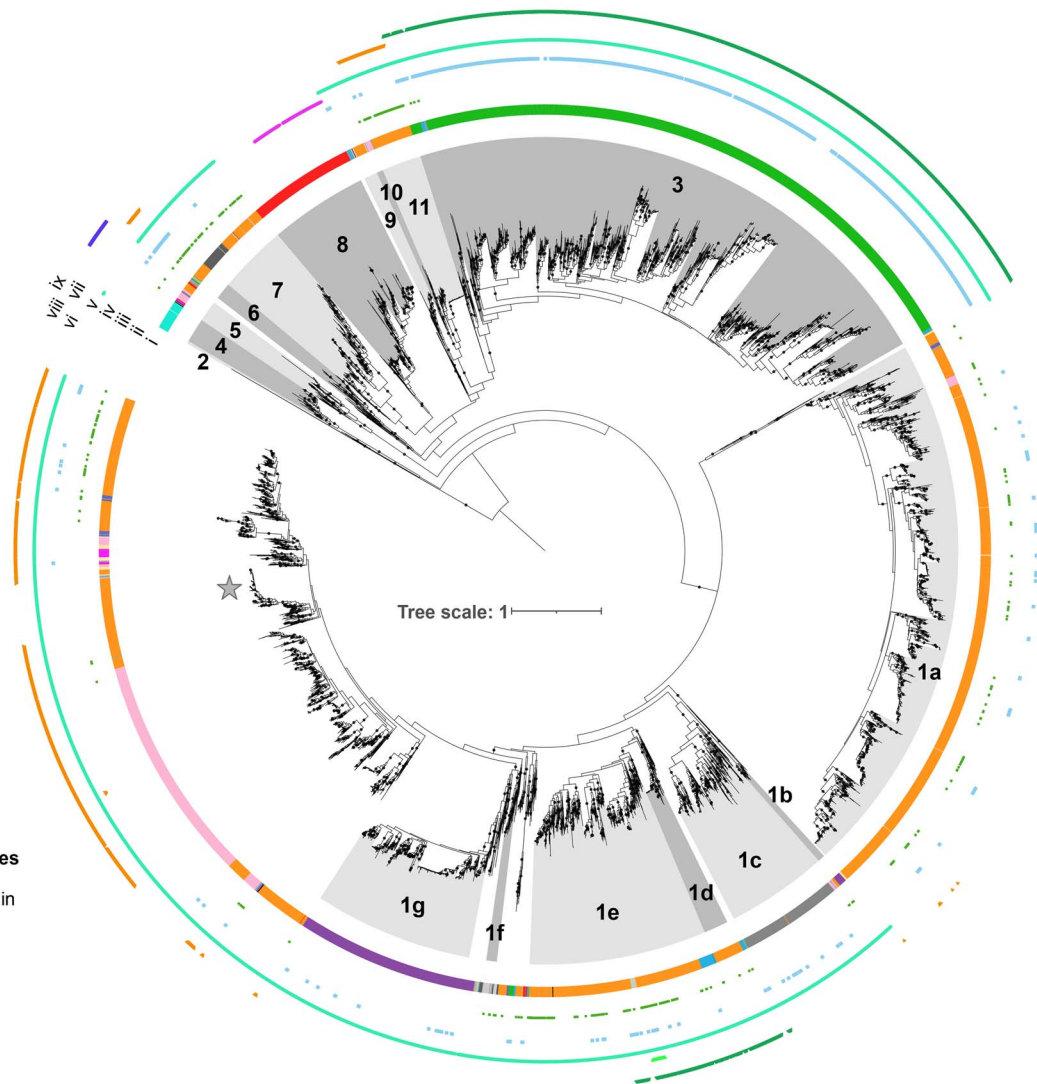
**Figure 3.** Maximum-likelihood NirK phylogeny. The phylogeny was inferred from 6,422 full-length NirK amino acid sequences using the LG+F+R10 model in IQ-TREE. Black circles on tree indicate nodes with SH-aLRT ≥ 80% and UFboot ≥ 95% support. Tips of tree are shaded according to clade, and data rings denote (i) phylum determined using GTDB-TK; (ii) presence of secondary *nirK* copies from other clades; (iii–ix) select NirK structural features. The outgroup, consisting of 367 multicopper oxidases derived from Cyanobacteria and Thermoprotetota genomes, is collapsed for clarity; clades are numbered and lettered following Wei *et al.* [29] and Bonilla-Rosso *et al.* [28] as much as possible, starting with the "canonical" Clade 1a; the star marks the location of the AniA NirK proteins from *N. gonorrhoeae*. [48]

bacteria within Deinococcota, Bacteroidota, Campylobacterota, Aquificota, and others (Fig. 2), and it was most prevalent in sewage (4%) and vents (5%; Fig. S3). Clade 1h consisted solely of anammox assemblies from a combination of environments. Although this clade was present at a median abundance of <1% in all biomes, it made up to 39% of reads in seafloor metagenomes, 25% in low latitude oceans, and 20% in rivers. It was also among the more negatively correlated groups with higher oxygen concentrations in ocean waters (Fig. 5A). Clade 2, which consisted of sequences derived from methane-oxidizing Proteobacteria, was often found in organisms with multiple copies of *nirS* and/or both *nirK* and *nirS* (Fig. 2 and Table S4). It accounted for a small but visible proportion of reads and surpassed that of the other methane-oxidizing clade, Clade 1i, which was dominated by sequences derived from Methylomirabilota. Clades 1d, 1e, and 4 were taxonomically diverse but were not prevalent in any biome (Fig. 2 and Fig. S3). Consistent with previous observations, Chloroflexi sequences

were polyphyletic [70]. The recently described Chloroflexi *nirS* gene fusion with a cytochrome c551/552 and cytochrome C mono/diheme variant [70] fell into Clade 1g and was rarely observed in the metagenomes. On the other hand, Clade 5, which consisted of sequences derived from a mixture of Chloroflexi, Proteobacteria, Actinobacteriota, and Planctomycetota MAGs and isolates from wastewater and thermophilic environments, dominated all terrestrial biomes except for croplands (22%–38% of sequences; Fig. S3).

Environmental distribution and functional potential of NirK clades

Similar to NirS, phylogenetic Mantel tests indicated that the overall functional dissimilarity based on KEGG orthologues and pathways correlated with the NirK sequence dissimilarity ($r = 0.388$ and $r = 0.359$, respectively; $P < 0.001$). ConsenTRAIT analysis found that hydrogen, methane, and manganese oxidation,
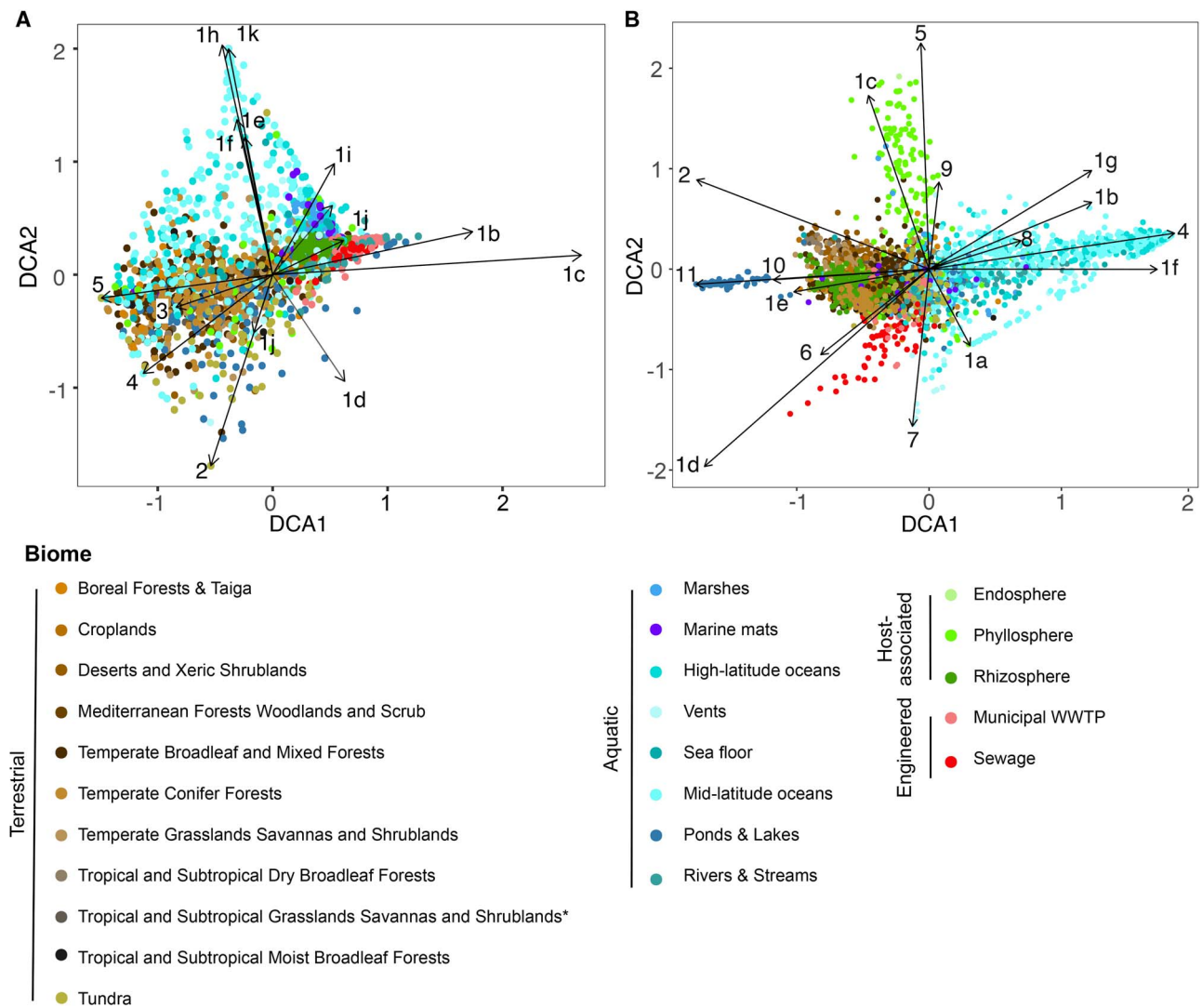
Figure 4. Nitrite reductase composition of globally distributed metagenomes. Ordinations of A) *nirS* and B) *nirK* clade composition based on detrended correspondence analysis. Each point represents the average of 100 rarefactions of 15 placements for a sample. Arrows terminate at the species score for each clade. Biomes represented by fewer than 10 samples after rarefying to 15 placements are excluded from the ordination.

arsenate reduction, and quinol-dependent nitric oxide reduction clustered on the NirK phylogeny (Table S7). Broad differences in *nirK* community composition, particularly between terrestrial and plant versus aquatic environments, were similarly apparent (Fig. 4B).

Specific NirK clades were often characterized by distinct functional genes and were associated with specific environments (Tables S10 and S11; Fig. 5C and D; Fig. S4). The "canonical" proteobacteria-dominated Clade 1a represented a variable proportion of the *nirK* sequences in all biomes (Fig. S4), although it was particularly abundant in drinking water treatment facilities. Subclades of Clade 1a were found at higher relative abundance in nitrate-rich and oxygen-poor ocean waters (Fig. 5C). The remainder of the NirK phylogeny consisted of clades representing all three domains of life, with strong biome-specific associations for some and a more cosmopolitan distribution for others. Among the clades with strong biome-specific associations were Clades 1c, 1d, 5, and 11. Burkholderia Clade 11 was particularly prevalent in ponds and lakes (59% median abundance), and it was strongly associated with genes for nitrous oxide reduction (i.e. *nosZ*) and complete denitrification compared to other clades

(Fig. 6C). Clade 1d housed the second genomic copy of *nirK* in Betaproteobacteria genomes and accounted for up to 15% of sequences in sewage. However, this clade rarely occurred in genomes encoding *nosZ* (Fig. 6C and Fig. S4). Clade 5 consisted of sequences derived from a mixture of taxonomically diverse alkaliphilic and thermophilic bacterial genera, including *Thioalkalivibrio, Caldalkalibacillus,* and *Thermus,* the genomes of which frequently encoded the ability for dissimilatory nitrate reduction to ammonium via the formate-dependent nitrite reductase. Consistent with the tolerance to extremophilic conditions known for these organisms, this clade was prevalent in hot springs metagenomes and was also common in the phyllosphere (Fig. S4). Clade 1c was dominated by halophilic archaea with small subclades of Proteobacteria and Bacteroidota (Fig. 3), and it was prevalent in the phyllosphere and in marshes (Fig. S4). Organisms within Clade 1c were often complete denitrifiers via quinol- or cytochrome c-dependent nitric oxide reductases [71] and the halophile-type nitrous oxide reductase (Fig. 6C; Supplementary Data 3; [72, 73]).

Clades 1e and 3 could be considered as "generalist" clades. Clade 1e consisted of sequences related to the hexameric
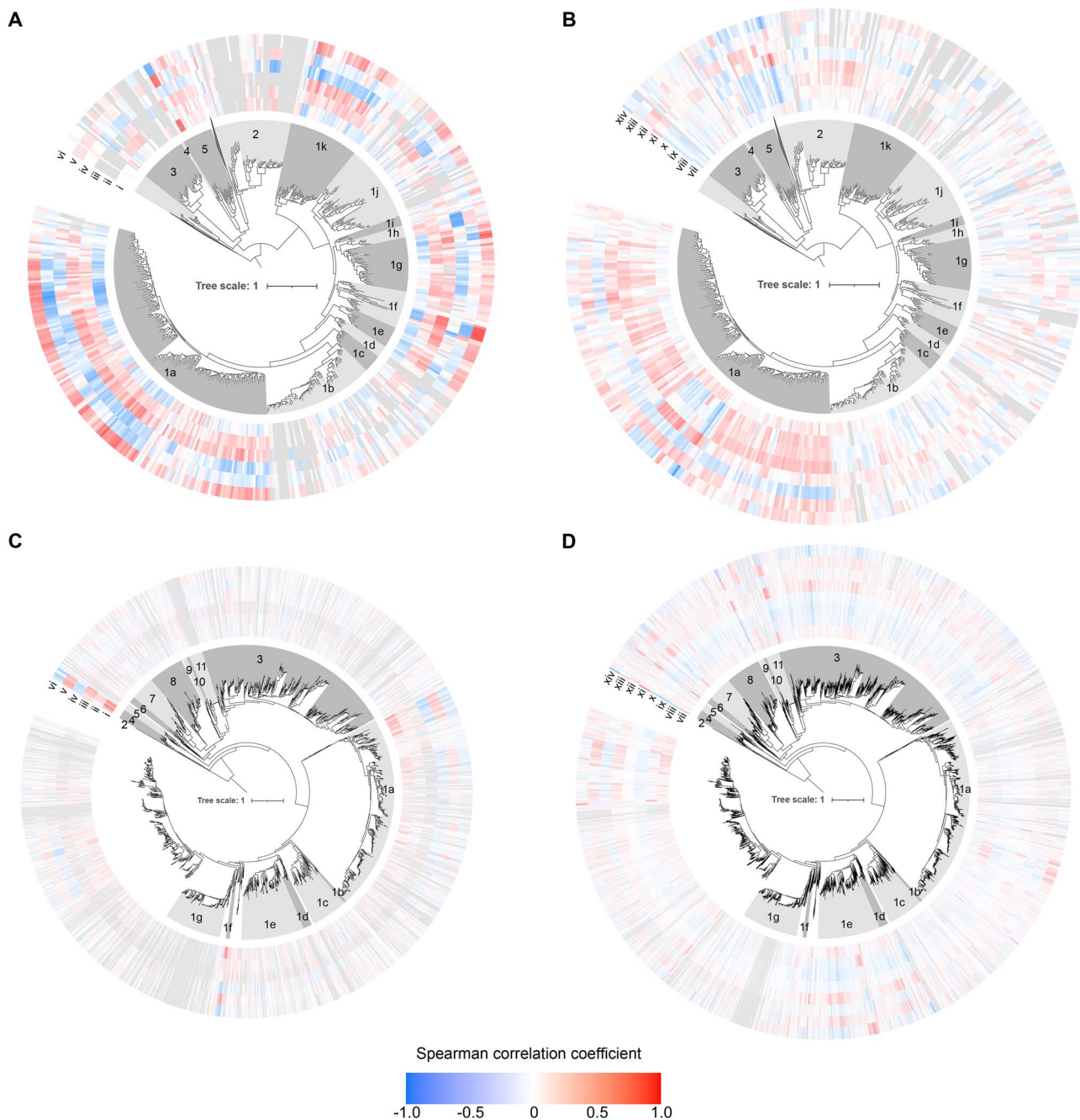
Figure 5. Environmental correlations of nitrite reductases and their clades. Spearman correlation of edge masses on A, B) NirS and C, D NirK phylogenies against environmental variables in A, C) water column seawater and B, D) terrestrial biomes. Values represent correlations in leaves of tree. Outer rings are coloured according to correlation strength, with grey indicating insufficient data. (i) water nitrate + nitrite, (ii) water ammonium, (iii) chlorophyll, (iv) dissolved oxygen, (v) water depth, (vi) water temperature, (vii) soil nitrate, (viii) soil ammonium, (ix) percent SOC, (x) soil moisture, (xi) pH in CaCl2, (xii) sand content, (xiii) copper, and (xiv) iron. These variables were selected for having data available from many metagenomes and were potentially relevant for structuring microbial communities (reactive N availability (i, ii, vii, viii), C availability or input (iii, ix), oxygenation degree (iii, iv, x), and physiochemical properties (v-vi, xi-xiv). Bootstrap support values have been removed for visual clarity. Metagenomes with >15 placements for the gene were included. Since not all metadata are available for all samples, correlations correspond to different subsets of samples [48].

Hyphomicrobium denitrificans and *Bradyrhizobium sp.* ORS 375 NirK [74, 75], and it was abundant across all terrestrial biomes (Fig. S4). This clade also contained the highest concentration of Chloroflexi sequences on the tree, driving the overall high abundance of this clade in wastewater treatment plants (Fig. S4). The large Clade 3 consisted of most sequences derived from Actinobacteriota, and it was prevalent across all biomes except for geysers (Fig. S4). Organisms carrying Clade 3 rarely encoded the nitrous oxide reductase NosZ, although a small fraction might

perform dissimilatory nitrate reduction to ammonium (Fig. 6B). The prevalence of this clade was generally positively correlated with soil pH or copper content, yet negatively correlated with soil organic matter concentration (Fig. 5D).

Nitrifiers fell into six clades in the NirK phylogeny (Figs 3, 7 and Table S6). Clade 1f consisted exclusively of sequences derived from the bacterial nitrite oxidizers and showed strong positive correlations with nitrate and nitrite availability and negative correlations with dissolved oxygen content (Fig. 5C). Clade 7
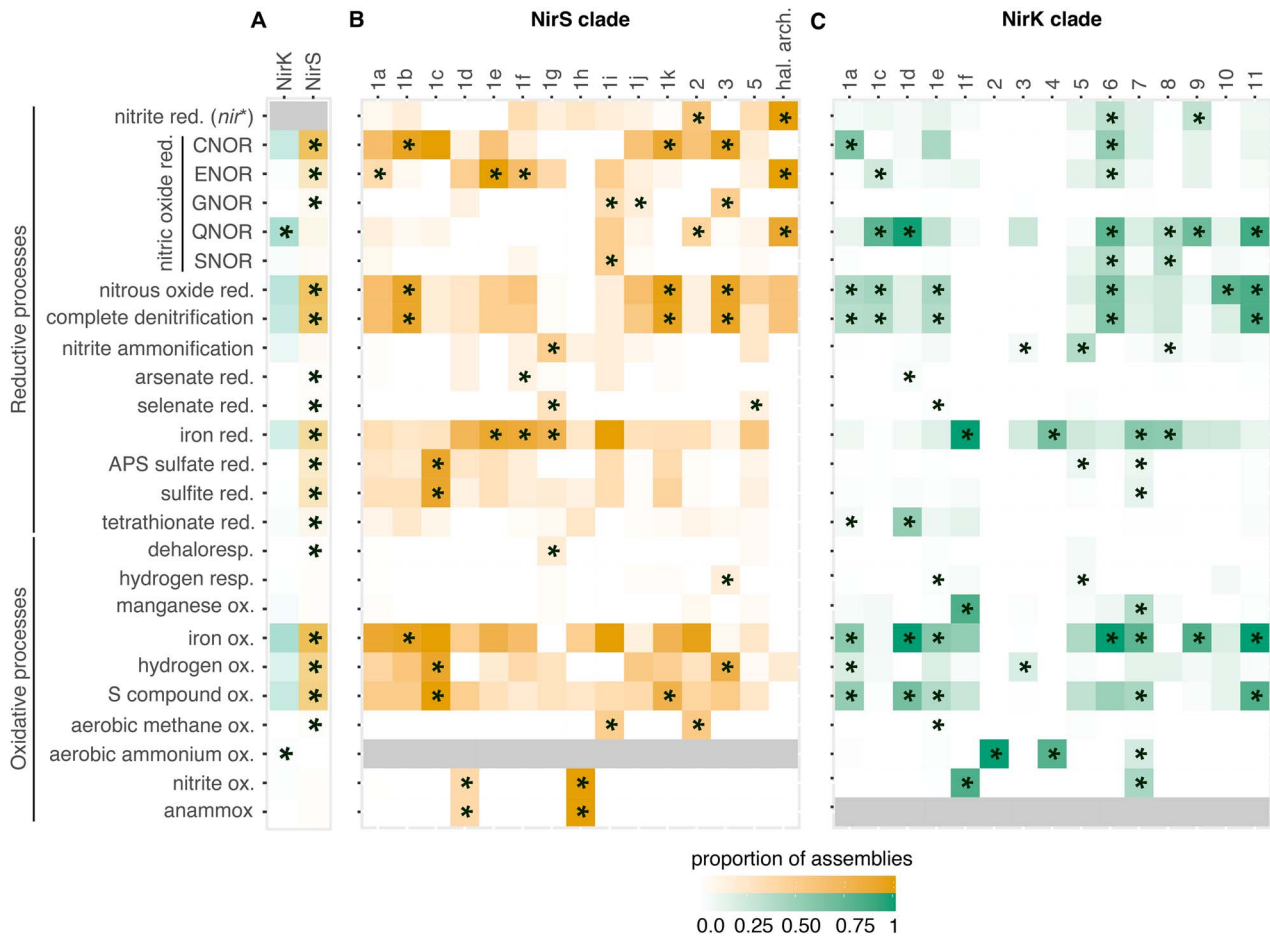
Figure 6. Proportion of assemblies carrying genes for redox traits. A) Proportion for all assemblies carrying the gene for NirS or NirK, or the proportion of assemblies within each clade carrying the genes for the associated function in the B) NirS and C) NirK phylogenies. Shading indicates the fraction of assemblies represented in the reference phylogeny which have the gene or trait of interest. Asterisks denote traits with a greater than expected frequency within a clade based on chi-squared post hoc tests. Grey indicates the trait was not tested due to not being found in members of any clade in the corresponding phylogeny. NirK clades 1b and 1g are not shown because they consist exclusively of eukaryotes and lack reliable protein predictions, while NirS clade 4 is excluded because it consists of just two taxa. *nir** refers to *nirS* when discussing assemblies carrying NirK and *nirK* when discussing assemblies carrying NirS. Assemblies carrying both *nirK* and *nirS* genes were excluded from panel A and are shown in Table S8. Genes used as markers for each trait are listed in the Supplementary Methods.

included all the complete ammonia oxidizers (comammox) and non-nitrifying Proteobacteria, with a handful of ammonia oxidizers. While Clades 7 and 4 dominated various aquatic biomes, Clade 2 was particularly abundant in dry terrestrial biomes and in dry, low SOM, high pH soils (Fig. 5C, D and Fig. S4). This propensity for Clade 2 to be found in dry soils was interesting, given that Nitrososphaera [76] ammonia oxidizing archaea have been reported to be more sensitive to drought than bacterial ammonia oxidizers [77, 78]. By contrast, NirK sequences from anammox bacteria were distributed in various regions of the phylogeny outside of these clades (Fig. 3).

Eukaryotic sequences were polyphyletic in the NirK tree [34, 79], with sequences from anoxic sediment- and oxygen minimum zone-dwelling foraminifera forming Clade 1b and with sequences from fungal denitrifiers clustering in Clade 1g (Fig. 3). The remaining eukaryotic sequences were derived from a combination of protists and algae, such as *Naegleria fowleri*, *Spumella vulgaris*, *Galderia sulphuraria*, and *Emiliania huxleyi*, and clustered with Clade 1g. While Clade 1g comprised a median of just 0.1% of the soil *nirK* reads and was not detected in 47% of metagenomes, it accounted for up to 26% of the sequences in ocean waters (median 0.8%; Fig. S4). By contrast, Clade 1b was uniformly rare (Fig. S4).

Most sequences which fell outside our defined clades (ca. 25%) occurred within a taxonomically and structurally diverse group represented in part by the *Neisseria gonorrhoea* AniA (Fig. 3). This group also included sequences from Bacteroidota, many of which are complete denitrifiers and were dominant in sewage and geyser metagenomes, but they were present at low abundance in ocean surface water samples, accounting for a median of 57% and 41% of *nirK* reads in the former and 1% in the latter. Among soils, it was most abundant in tundra (15%) and croplands (12%).

## Comparison of NirS and NirK ecology

Based on the redox traits associated with genomes encoding NirS or NirK as well as the biome distribution of *nirS* and *nirK* genes, potential differences in the ecology of organisms became apparent (Figs 6–8). Genomes carrying only *nirS* compared to those with only *nirK* were generally more likely to have all redox traits considered, except for aerobic ammonium oxidation and the quinol-dependent nitric oxide reductase (Fig. 6A). Cross-biome differences in the overall relative abundance of the two nitrite reductases showed that both *nir* genes were highly prevalent in engineered habitats and were rare on leaves and in the animal gut (Fig. 7A). However, *nirK* and *nirS* counts showed divergent
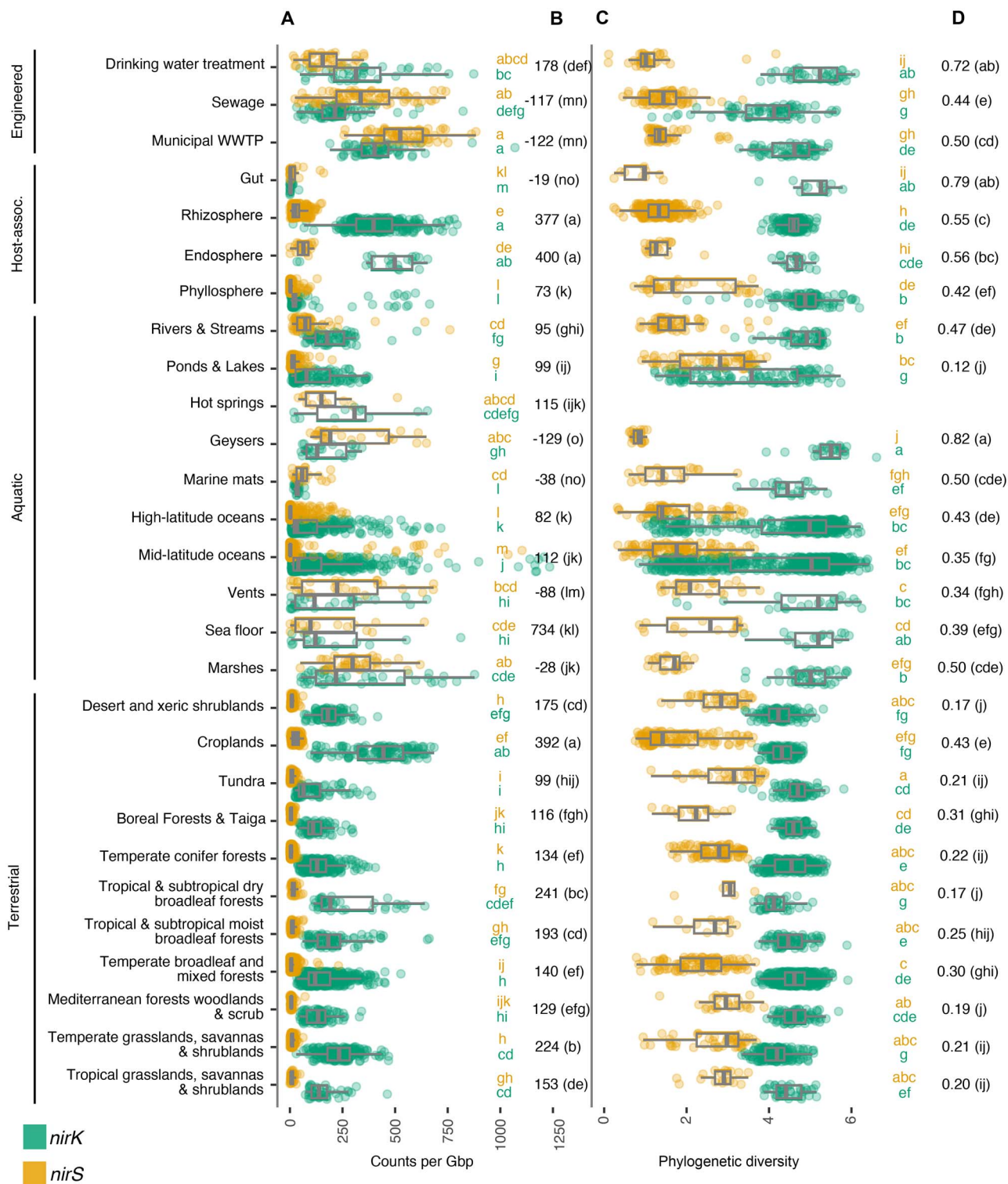
Figure 7. Abundance and diversity of *nir* genes in metagenomes across biomes. A) Abundance of *nirK* and *nirS* genes per gigabase (Gb) sequenced, with B) median difference in *nirK* and *nirS* abundance (δ*nir*). C) Diversity of *nirK* and *nirS* genes based on balance-weighted phylogenetic diversity (BWPD), with D) median ratios of *nirK:nirS* diversity. Ecosystem classifications are shown to the left. Biomes were compared using Benjamini-Hochberg FDR-corrected pairwise ranked comparisons following Kruskall-Wallis. Shared letters denoting similar abundance (in A) or diversity (in C) between biomes are shaded by *nir* type. Biomes represented by fewer than 10 metagenomes were excluded from the figure. Metagenomes off the scale of the axis are not shown but included in calculations for boxplot and pairwise comparisons. Boxplots show median and quartiles, and whiskers show 95 percentiles, and values for individual samples are shown as points.

biome associations, with *nirK* being more abundant than *nirS* in most biomes. Indeed, median *nirK* abundance was 19 times greater than that of *nirS* per Gb sequenced across all terrestrial

biomes, and it was particularly dominant in croplands and the rhizosphere. By contrast, *nirS* was only slightly more abundant than *nirK* in engineered biomes, vents, sea floor, geysers, hot
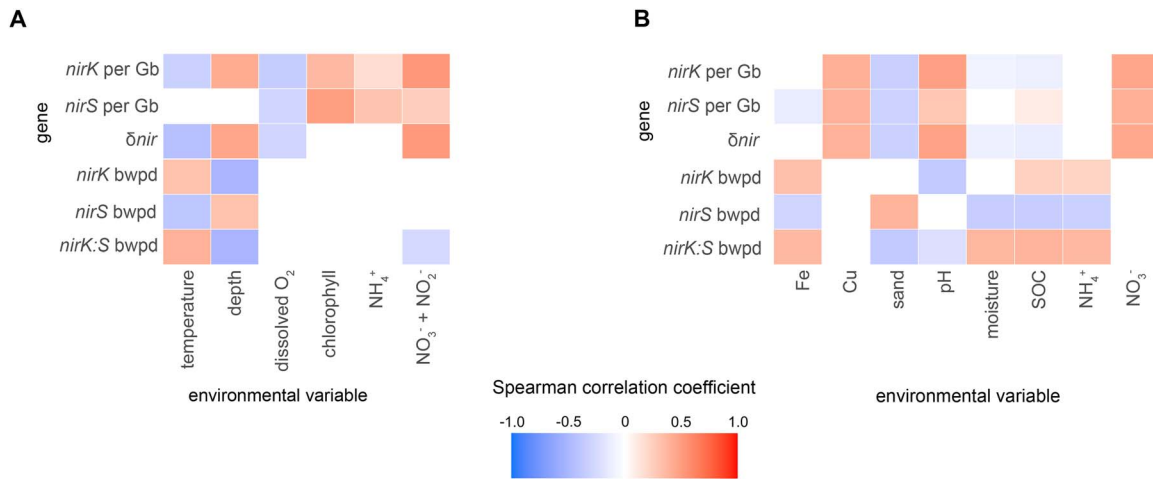
Figure 8. Environmental correlations of prevalence and diversity of nitrite reductase genes. Correlations between environmental variables and the abundance of *nirK* per Gb sequenced, *nirS* per Gb sequenced, ratio of their genes, alpha diversity (bwpd, balance-weighted phylogenetic diversity) and ratio of diversity, in A aquatic and B terrestrial metagenomes. SOC: soil organic carbon.

springs, and marshes and dominated most over *nirK* in geysers and wastewater (Fig. 7B). In pelagic marine samples, *nirS* was 34 times less prevalent than *nirK*. The levels of nitrate or nitrite were positively correlated with *nirK* and *nirS* counts and the prevalence of *nirK* over *nirS* (*δnir*) in both soils and marine samples (Fig. 8; Tables S12 and S13). Variables indicative of environment redox state, like dissolved oxygen in marine samples and soil moisture in soil samples, were negatively correlated with the abundance of both nitrite reductases, and soil moisture indicated the prevalence of *nirS* over *nirK* (negative *δnir*) at lower redox potentials. Yet, genome comparison indicated that organisms carrying both *nir* genes may be better adapted to low redox conditions than those encoding just one or the other, considering their genomes are also more likely to encode the potential for most redox traits examined (Table S14).

BWPD [67] was also biome-specific for the two *nir* genes (Fig. 7C and D). Metagenomes from drinking water treatment and geysers were characterized by high *nirK* but low *nirS* diversity resulting in a high *nirK*:*nirS* diversity ratio, while ponds were characterized by among the highest *nirS* but lowest *nirK* diversity and thus a low *nirK*:*nirS* diversity ratio. Abundance and diversity were negatively correlated for both *nirK* and *nirS* ($r = -0.45$ and -0.57, respectively; $P < 0.001$), indicating that the environments where nitrite reduction capacity is most beneficial particularly benefits only a restricted part of the nitrite reductase phylogeny. For NirS, the decrease in diversity could largely be attributed to an enrichment of Clade 1k in aquatic systems, while for NirK, it depended on biome. Diversities of *nirS* and *nirK* tended to show contrasting correlations with the abiotic variables examined, including an increase in the *nirK*:*nirS* diversity ratio in soils characterized by high iron, SOC, and/or ammonium contents (Fig. 8).

Discussion

This study provides novel insights into the phylogenetic structure of two NO-forming nitrite reductases and their distribution across global biomes. We demonstrate that NirK and NirS are diverse proteins whose evolution is associated with both ecological niche (Figs 4 and 6) and organismal phylogeny (Figs 2 and 3), and that genes encoding these two nitrite reductases have strong biome associations (Figs 5, 7, 8 and Figs S3, S4). We further show that

despite performing the same chemical reaction, organisms carrying multiple copies, clades, or even classes of nitrite reductases are not rare (Tables S4 and S6). Our results contest the original interpretation of the two kinds of nitrite reductase as functionally redundant [1]. Not only did we find that it is 5–10 times more common than previously thought for genomes to encode multiple nitrite reductases [2], but genes for the two variants also show divergent biome prevalence and correlations with environmental variables (Figs 7 and 8). Usually thought as exclusively associated with anaerobic respiration, nitrite reduction can also be involved in nitrite detoxification [22, 80] or simply support the elimination of excess reducing power to allow other physiological processes to continue [81]. Although the physiological benefits of producing multiple nitrite reductases is an area of active research [4, 21, 82], in principle, it could allow organisms to continue to denitrify across a wider range of environments or under rapidly shifting conditions. The observation that assemblies carrying both *nirS* and *nirK* were more likely to encode genes for many of the other traits examined supports the hypothesis that organisms carrying both have a greater degree of redox flexibility than assemblies carrying *nirK* or *nirS* alone, which should be addressed in future research.

Adaptation to shifting environmental conditions is evident in the conditions under which the two genes and their constituent clades were most prevalent. Biomes such as sewage, WWTP, and marshes were characterized by a high overall nitrite reductase abundance and were often dominated by *nirS*, and particularly the "canonical" Clade 1a, while soils and pelagic aquatic biomes tended to be dominated by *nirK*. Previous studies on the ecology of nitrite reductases often concluded that *nirS* abundance and/or diversity are more sensitive to changing environmental conditions than *nirK* [11, 13, 24, 83, 84]. This contrasts with our results showing similar or lower correlations between *nirS* abundance or diversity and environmental factors. Our results also contrast the observations that *nirK* diversity increases with soil pH [11] and *nirS* diversity increases with ammonium content [9]. These previous studies are largely primer-based and constrained to evaluating the response of Clade 1a of both genes, which our work demonstrates is particularly poorly representative of the *nirK*-type nitrite reductase community. However, consistent with our observations, a metagenome-based study found that wet, anoxic soils favour *nirS* over *nirK*, which the authors attributed to the greater

biosynthetic cost of the former [10]. The positive correlation with moisture was driven by a decrease in *nirK* prevalence rather than increased *nirS* in the present study, and it largely corresponded to a reduction in the diversity of *nirS* and a positive correlation with moisture for *nirS* Clade 1a. We propose that the reduction in *nirK* counts in moist soils may also be driven by the oxygen requirement of organisms with this Nir variant, including the ammonia-oxidizing archaea encoding NirK clade 2 as well as groups of Actinobacteriota or fungi which reduce nitrite at a similar or more rapid rate in the presence of oxygen [85, 86]. It is also possible that mechanisms driving the relative prevalence of *nirK* and *nirS* differ between the intercontinental scale of our analysis and the more localized scales of earlier studies.

By expanding the phylogeny to include diverse eukaryotic NirK sequences and extending the study into additional biomes, we confirmed that, contrary to primer-based approaches that also pick up a substantial number of bacterial *nirK* reads [87–89], fungal nitrite reductase sequences are rare in soil [90]. However, our study found them to be occasionally abundant in the marine water column. Given that known fungal denitrifiers have an oxygen requirement, the sporadic prevalence within pelagic samples may be attributed to the overrepresentation of samples derived from well-oxygenated surface waters within our database (90% of samples have >4 mg O_2/l). Overall, our analysis provides a global assessment and adds nuance to some of the previously reported drivers of NirK and NirS diversity and abundance, which were restricted to specific habitats or clades of the nitrite reductase phylogenies not necessarily representative of the full diversity of organisms carrying each gene.

Finally, our study addressed the co-occurrence of NirS and NirK subclades with other functional genes, such as those involved in iron and sulphur redox reactions. In this analysis, we did not apply characteristic motif verification [91] as we did for the Nir sequences, so the presence of these additional genes may be overestimated. Nonetheless, our analysis shows that organisms encoding NirS, as opposed to NirK, display the genetic potential for reducing and oxidizing a broad range of substrates, although this varies between clades. Of note is the potential for the further processing of nitric oxide into the greenhouse gas nitrous oxide by nitric oxide reductase and eventually to N_2 by NosZ. Previously, the cytochrome c-dependent nitric oxide reductase was proposed to be exclusively involved in denitrification, while the quinol-dependent variant may also detoxify nitric oxide [92]. Accordingly, assemblies carrying NirS more commonly encode the cytochrome c-dependent nitric oxide reductase and NosZ (i.e. potential for complete denitrification), and there is a much lower prevalence of *nosZ* among assemblies carrying *nirK*. This highlights the divergent roles that NirK and NirS may play in the environment. Yet, this broader pattern of reduced potential for complete denitrification among NirK microorganisms was largely driven by the absence or near absence of the complete denitrification trait among NirK Clades 2–4, represented by ammonia-oxidizing archaea and Actinobacteriota. This includes the absence of Nor in many cases, indicating that these organisms must either tolerate or process the potentially toxic NO in another way [27, 93]. The weak co-association between NirK and the cytochrome c-dependent nitric oxide reductase, as well as NosZ, combined with the much higher abundance of *nirK* than *nirS* across most biomes further indicate that Nir-driven NO-forming nitrite reduction is not solely associated with "canonical" denitrification, particularly in marine surface waters and dry soils. For example, it has been suggested that NirK in the bacterial ammonia oxidizer *Nitrosomonas europaea* may catalyze the oxidation of nitric oxide to nitrite [94]. This

suggests a cautious interpretation of denitrification capacity in count-based analyses of environmental metagenomes. Our work provides a framework beyond gene annotation that integrates phylogenetic background, protein structure, genomic context, and environmental distributions for analysing environmental *nir* genes based on clade-level information that can help infer the fate of nitrite and the processes to which its reduction is most likely to be linked to. Nitrite reduction is a complex ecological trait, and our research implies its understanding and effective management necessitates considering it as such.

Author contributions

Sara Hallin (Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing, Writing—review & editing), Germán Bonilla-Rosso (Conceptualization, Data curation, Investigation, Writing, Writing—review & editing), Aurélien Saghaï (Data curation, Investigation, Software, Visualization, Writing—review & editing), Christopher M. Jones (Formal analysis, Investigation, Software, Writing—review & editing), Grace Pold (Data curation, Formal analysis, Investigation, Software, Visualization, Writing, Writing—review & editing), and Marc Strous (Writing—review & editing).

Supplementary material

Supplementary material is available at *ISME Communications* online.

Data availability

All genomes and metagenomes used in this analysis are publicly available and documented with their sources in Tables S1, S4, and S6. Reference alignments, clade-specific HMMs, and phylogenies for each protein are available in figshare (doi:10.6084/m9.figshare.23913078).

References

1. Zumft WG. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* 1997;**61**:84.

2. Graf DRH, Jones CM, Hallin S. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. *PLoS One* 2014;**9**:e114118. https://doi.org/10.1371/journal.pone.0114118

3. Glockner AB, Jüngst A, Zumft WG. Copper-containing nitrite reductase from *Pseudomonas aureofaciens* is functional in a mutationally cytochrome cd_1-free background (NirS$^−$) of *Pseudomonas stutzeri*. *Arch Microbiol* 1993;**160**:18–26. https://doi.org/10.1007/bf00258141

4. Sánchez C, Minamisawa K. Redundant roles of *Bradyrhizobium oligotrophicum* Cu-type (NirK) and cd_1-type (NirS) nitrite reductase genes under denitrifying conditions. *FEMS Microbiol Lett* 2018;**365**:fny015. https://doi.org/10.1093/femsle/fny015

5. Tuomainen JM, Hietanen S, Kuparinen J *et al.* Baltic Sea cyanobacterial bloom contains denitrification and nitrification genes, but has negligible denitrification activity. *FEMS Microbiol Ecol* 2003;**45**:83–96. https://doi.org/10.1016/S0168-6496(03)00131-4

6. Abell G, Revill A, Smith C *et al.* Archaeal ammonia oxidizers and *nirS*-type denitrifiers dominate sediment nitrifying and denitrifying populations in a subtropical macrotidal estuary. *ISME J* 2009;**4**:286–300. https://doi.org/10.1038/ismej.2009.105

7. Philippot L, Cuhel J, Saby NPA *et al.* Mapping field-scale spatial patterns of size and activity of the denitrifier community. *Environ Microbiol* 2009;**11**:1518–26. https://doi.org/10.1111/j.1462-2920.2009.01879.x

8. Enwall K, Throbäck IN, Stenberg M *et al.* Soil resources influence spatial patterns of denitrifying communities at scales compatible with land management. *Appl Environ Microbiol* 2010;**76**:2243–50. https://doi.org/10.1128/AEM.02197-09

9. Jones CM, Hallin S. Ecological and evolutionary factors underlying global and local assembly of denitrifier communities. *ISME J* 2010;**4**:633–41. https://doi.org/10.1038/ismej.2009.152

10. Nadeau SA, Roco CA, Debenport SJ *et al.* Metagenomic analysis reveals distinct patterns of denitrification gene abundance across soil moisture, nitrate gradients. *Environ Microbiol* 2019;**21**:1255–66. https://doi.org/10.1111/1462-2920.14587

11. Kou Y, Liu Y, Li J *et al.* Patterns and drivers of *nirK*-type and *nirS*-type denitrifier community assembly along an elevation gradient. *mSystems* 2021;**6**:e0066721. https://doi.org/10.1128/mSystems.00667-21

12. Gowda K, Ping D, Mani M *et al.* Genomic structure predicts metabolite dynamics in microbial communities. *Cell* 2022;**185**:530–546.e25. https://doi.org/10.1016/j.cell.2021.12.036

13. Sun H, Jiang S. A review on *nirS*-type and *nirK*-type denitrifiers via a scientometric approach coupled with case studies. *Environ Sci Process Impacts* 2022;**24**:221–32. https://doi.org/10.1039/D1EM00518A

14. Kobayashi S, Hira D, Yoshida K *et al.* Nitric oxide production from nitrite reduction and hydroxylamine oxidation by copper-containing dissimilatory nitrite reductase (NirK) from the aerobic ammonia-oxidizing archaeon, *Nitrososphaera viennensis*. *Microbes Environ* 2018;**33**:428–34. https://doi.org/10.1264/jsme2.ME18058

15. Cantera JJL, Stein LY. Role of nitrite reductase in the ammonia-oxidizing pathway of *Nitrosomonas europaea*. *Arch Microbiol* 2007;**188**:349–54. https://doi.org/10.1007/s00203-007-0255-4

16. Harrold ZR, Skidmore ML, Hamilton TL *et al.* Aerobic and anaerobic thiosulfate oxidation by a cold-adapted, subglacial chemoautotroph. *Appl Environ Microbiol* 2016;**82**:1486–95. https://doi.org/10.1128/AEM.03398-15

17. Klueglein N, Kappler A. Abiotic oxidation of Fe(II) by reactive nitrogen species in cultures of the nitrate-reducing Fe(II) oxidizer *Acidovorax sp.* BoFeN1—questioning the existence of enzymatic Fe(II) oxidation. *Geobiology* 2013;**11**:180–90. https://doi.org/10.1111/gbi.12019

18. Su JF, Zheng SC, Huang TL *et al.* Characterization of the anaerobic denitrification bacterium *Acinetobacter sp.* SZ28 and its application for groundwater treatment. *Bioresour Technol* 2015;**192**:654–9. https://doi.org/10.1016/j.biortech.2015.06.020

19. van de Vossenberg J, Woebken D, Maalcke WJ *et al.* The metagenome of the marine anammox bacterium 'Candidatus *Scalindua profunda*' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ Microbiol* 2013;**15**:1275–89. https://doi.org/10.1111/j.1462-2920.2012.02774.x

20. Strous M, Pelletier E, Mangenot S *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 2006;**440**:790–4. https://doi.org/10.1038/nature04647

21. Wittorf L, Jones CM, Bonilla-Rosso G *et al.* Expression of *nirK* and *nirS* genes in two strains of *Pseudomonas stutzeri* harbouring both types of NO-forming nitrite reductases. *Res Microbiol* 2018;**169**:343–7. https://doi.org/10.1016/j.resmic.2018.04.010

22. Beaumont HJE, Hommes NG, Sayavedra-Soto LA *et al.* Nitrite reductase of *Nitrosomonas europaea* is not essential for production of gaseous nitrogen oxides and confers tolerance to nitrite. *J Bacteriol* 2002;**184**:2557–60. https://doi.org/10.1128/JB.184.9.2557-2560.2002

23. Basaglia M, Toffanin A, Baldan E *et al.* Selenite-reducing capacity of the copper-containing nitrite reductase of *Rhizobium sullae*. *FEMS Microbiol Lett* 2007;**269**:124–30. https://doi.org/10.1111/j.1574-6968.2006.00617.x

24. Abada A, Beiralas R, Narvaez D *et al.* Aerobic bacteria produce nitric oxide via denitrification and promote algal population collapse. *ISME J* 2023;**17**:1167–83. https://doi.org/10.1038/s41396-023-01427-8

25. Yamazaki T, Oyanagi H, Fujiwara T *et al.* Nitrite reductase from the magnetotactic bacterium *Magnetospirillum magnetotacticum*. A novel cytochrome cd_1 with Fe(II):nitrite oxidoreductase activity. *Eur J Biochem* 1995;**233**:665–71. https://doi.org/10.1111/j.1432-1033.1995.665_2.x

26. Li Y, Bali S, Borg S *et al.* Cytochrome cd_1 nitrite reductase NirS is involved in anaerobic magnetite biomineralization in *Magnetospirillum gryphiswaldense* and requires NirN for proper d1 heme assembly. *J Bacteriol* 2013;**195**:4297–309. https://doi.org/10.1128/JB.00686-13

27. Ettwig KF, Butler MK, Le Paslier D *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 2010;**464**:543–8. https://doi.org/10.1038/nature08883

28. Bonilla-Rosso G, Wittorf L, Jones CM *et al.* Design and evaluation of primers targeting genes encoding NO-forming nitrite reductases: implications for ecological inference of denitrifying communities. *Sci Rep* 2016;**6**:39208. https://doi.org/10.1038/srep39208

29. Wei W, Isobe K, Nishizawa T *et al.* Higher diversity and abundance of denitrifying microorganisms in environments than considered previously. *ISME J* 2015;**9**:1954–65. https://doi.org/10.1038/ismej.2015.9

30. Jones CM, Stres B, Rosenquist M *et al.* Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification. *Mol Biol Evol* 2008;**25**:1955–66. https://doi.org/10.1093/molbev/msn146

31. Shapiro BJ, Polz MF. Microbial speciation. *Cold Spring Harb Perspect Biol* 2015;**7**:a018143. https://doi.org/10.1101/cshperspect.a018143

32. Heylen K, Gevers D, Vanparys B *et al.* The incidence of *nir*S and *nir*K and their genetic heterogeneity in cultivated denitrifiers. *Environ Microbiol* 2006;**8**:2012–21. https://doi.org/10.1111/j.1462-2920.2006.01081.x

33. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195. https://doi.org/10.1371/journal.pcbi.1002195

34. Woehle C, Roy A-S, Glock N *et al.* Denitrification in foraminifera has an ancient origin and is complemented by associated bacteria. *Proc Natl Acad Sci* 2022;**119**:e2200198119. https://doi.org/10.1073/pnas.2200198119

35. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* 2001;**17**:282–3. https://doi.org/10.1093/bioinformatics/17.3.282

36. Li W, Godzik A. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;**22**:1658–9. https://doi.org/10.1093/bioinformatics/btl158

37. 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

38. Manni M, Berkeley MR, Seppey M *et al.* BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 2021;**38**:4647–54. https://doi.org/10.1093/molbev/msab199

39. Mirdita M, Steinegger M, Breitwieser F *et al.* Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* 2021;**37**:3029–31. https://doi.org/10.1093/bioinformatics/btab184

40. Suzek BE, Wang Y, Huang H *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 2015;**31**:926–32. https://doi.org/10.1093/bioinformatics/btu739

41. Chaumeil P-A, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 2020;**36**:1925–7. https://doi.org/10.1093/bioinformatics/btz848

42. Nurizzo D, Silvestrini MC, Mathieu M *et al.* N-terminal arm exchange is observed in the 2.15 A crystal structure of oxidized nitrite reductase from Pseudomonas aeruginosa. *Structure* 1997;**5**:1157–71. https://doi.org/10.1016/S0969-2126(97)00267-0

43. Cutruzzolà F, Brown K, Wilson EK *et al.* The nitrite reductase from *Pseudomonas aeruginosa*: essential role of two active-site histidines in the catalytic and structural properties. *Proc Natl Acad Sci* 2001;**98**:2232–7. https://doi.org/10.1073/pnas.041365298

44. Bartossek R, Nicol GW, Lanzen A *et al.* Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* 2010;**12**:1075–88. https://doi.org/10.1111/j.1462-2920.2010.02153.x

45. Westram R, Bader K, Preusse E *et al.* ARB: a software environment for sequence data. In: Frans J. de Bruijn. (eds.), *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*. Wiley-Blackwell, Hoboken, NJ, 2011;399–406.

46. Minh BQ, Schmidt HA, Chernomor O *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

47. Guindon S, Dufayard J-F, Lefort V *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21. https://doi.org/10.1093/sysbio/syq010

48. 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

49. Hug LA, Baker BJ, Anantharaman K *et al.* A new view of the tree of life. *Nat Microbiol* 2016;**1**:1–6. https://doi.org/10.1038/nmicrobiol.2016.48

50. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019;**35**:4162–4. https://doi.org/10.1093/bioinformatics/btz188

51. Boyd JA, Woodcroft BJ, Tyson GW. *Comparative Genomics Using EnrichM*, 2019. https://github.com/geronimp/enrichM.

52. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;**28**:27–30. https://doi.org/10.1093/nar/28.1.27

53. Harmon LJ, Glor RE. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* 2010;**64**:2173–8. https://doi.org/10.1111/j.1558-5646.2010.00973.x

54. Martiny AC, Treseder K, Pusch G. Phylogenetic conservatism of functional traits in microorganisms. *ISME J* 2013;**7**:830–8. https://doi.org/10.1038/ismej.2012.160

55. Glass GV, McGaw B, Smith ML. *Meta-Analysis in Social Research*. SAGE, Beverley Hills, CA, 1981.

56. R Core Team. *R: A Language and Environment for Statistical Computing*, 2023. https://www.r-project.org/.

57. Olson DM, Dinerstein E. The global 200: priority ecoregions for global conservation. *Ann Mo Bot Gard* 2002;**89**:199–224. https://doi.org/10.2307/3298564

58. Pebesma E, Bivand R, Rowlingson B *et al. sp: Classes and Methods for Spatial Data, R News* 2005;**5**:9–13. https://geobgu.xyz/r-2019/resources/Rnews_2005-2.pdf

59. Bivand R, Keitt T, Rowlingson B *et al. Rgdal: Bindings for the "Geospatial" Data Abstraction Library*, 2022. http://rgdal.r-forge.r-project.org, https://gdal.org, https://proj.org, https://r-forge.r-project.org/projects/rgdal/

60. Bivand R, Rundel C, Pebesma E *et al. Rgeos: Interface to Geometry Engine—Open Source ('GEOS')*, 2021. https://r-forge.r-project.org/projects/rgeos/, https://libgeos.orghttp://rgeos.r-forge.r-project.org/index.html

61. Boyd JA, Woodcroft BJ, Tyson GW. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res* 2018;**46**:e59. https://doi.org/10.1093/nar/gky174

62. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3. https://doi.org/10.1093/bioinformatics/btu033

63. Rambaut A. FigTree v.1.4.4-1. 2018. http://tree.bio.ed.ac.uk/software/figtree/

64. Nayfach S, Roux S, Seshadri R *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 2021;**39**:499–509. https://doi.org/10.1038/s41587-020-0718-6

65. Czech L, Barbera P, Stamatakis A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* 2020;**36**:3263–5. https://doi.org/10.1093/bioinformatics/btaa070

66. Matsen FA, Kodner RB, Armbrust EV. Pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 2010;**11**:538. https://doi.org/10.1186/1471-2105-11-538

67. McCoy CO, Matsen FA. Abundance-weighted phylogenetic diversity measures distinguish microbial community states and

are robust to sampling depth. *PeerJ* 2013;**1**:e157. https://doi.org/10.7717/peerj.157

68. Oksanen J, Simpson GL, Blanchet FG *et al*. Vegan: Community Ecology Package, 2022. https://cran.r-project.org/web/packages/vegan/vegan.pdf

69. Zajicek RS, Bali S, Arnold S *et al*. d_1 haem biogenesis—assessing the roles of three *nir* gene products. *FEBS J* 2009;**276**:6399–411. https://doi.org/10.1111/j.1742-4658.2009.07354.x

70. Schwartz SL, Momper L, Rangel LT *et al*. Novel nitrite reductase domain structure suggests a chimeric denitrification repertoire in the phylum Chloroflexi. *MicrobiologyOpen* 2022;**11**:e1258. https://doi.org/10.1002/mbo3.1258

71. Torregrosa-Crespo J, González-Torres P, Bautista V *et al*. Analysis of multiple haloarchaeal genomes suggests that the quinone-dependent respiratory nitric oxide reductase is an important source of nitrous oxide in hypersaline environments. *Environ Microbiol Rep* 2017;**9**:788–96. https://doi.org/10.1111/1758-2229.12596

72. Sanford RA, Wagner DD, Wu Q *et al*. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc Natl Acad Sci* 2012;**109**:19709–14. https://doi.org/10.1073/pnas.1211238109

73. Jones CM, Graf DR, Bru D *et al*. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. *ISME J* 2013;**7**:417–26. https://doi.org/10.1038/ismej.2012.125

74. Nojiri M, Xie Y, Inoue T *et al*. Structure and function of a hexameric copper-containing nitrite reductase. *Proc Natl Acad Sci* 2007;**104**:4315–20. https://doi.org/10.1073/pnas.0609195104

75. Sasaki D, Watanabe TF, Eady RR *et al*. Reverse protein engineering of a novel 4-domain copper nitrite reductase reveals functional regulation by protein–protein interaction. *FEBS J* 2021;**288**:262–80. https://doi.org/10.1111/febs.15324

76. Séneca J, Pjevac P, Canarini A *et al*. Composition and activity of nitrifier communities in soil are unresponsive to elevated temperature and CO2, but strongly affected by drought. *ISME J* 2020;**14**:3038–53. https://doi.org/10.1038/s41396-020-00735-7

77. Thion C, Prosser JI. Differential response of nonadapted ammonia-oxidising archaea and bacteria to drying–rewetting stress. *FEMS Microbiol Ecol* 2014;**90**:380–9. https://doi.org/10.1111/1574-6941.12395

78. Bello MO, Thion C, Gubry-Rangin C *et al*. Differential sensitivity of ammonia oxidising archaea and bacteria to matric and osmotic potential. *Soil Biol Biochem* 2019;**129**:184–90. https://doi.org/10.1016/j.soilbio.2018.11.017

79. Woehle C, Roy A-S, Glock N *et al*. A novel eukaryotic denitrification pathway in foraminifera. *Curr Biol* 2018;**28**:2536–2543.e5. https://doi.org/10.1016/j.cub.2018.06.027

80. Stein LY, Arp DJ. Loss of ammonia monooxygenase activity in *Nitrosomonas europaea* upon exposure to nitrite. *Appl Environ Microbiol* 1998;**64**:4098–102. https://doi.org/10.1128/AEM.64.10.4098-4102.1998

81. Torregrosa-Crespo J, Esclapez J, Bautista V *et al*. Denitrification in extreme environments. In: Durvasula R., Rao D.V.S. (eds.), *Extremophiles: From Biology to Biotechnology*. CRC Press, Boca Raton, 2018.

82. Jang J, Ashida N, Kai A *et al*. Presence of Cu-type (NirK) and cd_1-type (NirS) nitrite reductase genes in the denitrifying bacterium *Bradyrhizobium nitroreducens* sp. nov. *Microbes Environ* 2018;**33**:326–31. https://doi.org/10.1264/jsme2.ME18039

83. Chen Z, Ma J, Liu Y *et al*. Differential responses of soil *nirS*- and *nirK*-type denitrifying microbial communities to long-term application of biogas slurry in a paddy soil. *Appl Soil Ecol* 2023;**182**:104711. https://doi.org/10.1016/j.apsoil.2022.104711

84. Li M, Hong Y, Cao H *et al*. Diversity, abundance, and distribution of NO-forming nitrite reductase-encoding genes in deep-sea subsurface sediments of the South China Sea. *Geobiology* 2013;**11**:170–9. https://doi.org/10.1111/gbi.12020

85. Zhou Z, Takaya N, Sakairi MAC *et al*. Oxygen requirement for denitrification by the fungus *Fusarium oxysporum*. *Arch Microbiol* 2001;**175**:19–25. https://doi.org/10.1007/s002030000231

86. Shoun H, Kano M, Baba I *et al*. Denitrification by Actinomycetes and purification of dissimilatory nitrite reductase and azurin from *Streptomyces thioluteus*. *J Bacteriol* 1998;**180**:4413–5. https://doi.org/10.1128/JB.180.17.4413-4415.1998

87. Ma Y, Zilles JL, Kent AD. An evaluation of primers for detecting denitrifiers via their functional genes. *Environ Microbiol* 2019;**21**:1196–210. https://doi.org/10.1111/1462-2920.14555

88. Bösch Y, Jones CM, Finlay R *et al*. Minimizing tillage modifies fungal denitrifier communities, increases denitrification rates and enhances the genetic potential for fungal, relative to bacterial, denitrification. *Soil Biol Biochem* 2022;**170**:108718. https://doi.org/10.1016/j.soilbio.2022.108718

89. Lourenço KS, Costa OY D A, Cantarella H *et al*. Ammonia-oxidizing bacteria and fungal denitrifier diversity are associated with N2O production in tropical soils. *Soil Biol Biochem* 2022;**166**:108563. https://doi.org/10.1016/j.soilbio.2022.108563

90. Bösch Y, Pold G, Saghaï A *et al*. Distribution and environmental drivers of fungal denitrifiers in global soils. *Microbiol Spectr* 2023;**0**:e00061–23

91. Zhou Z, Tran PQ, Breister AM *et al*. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* 2022;**10**:33. https://doi.org/10.1186/s40168-021-01213-8

92. Hendriks J, Oubrie A, Castresana J *et al*. Nitric oxide reductases in bacteria. *Biochim Biophys Acta BBA—Bioenerg* 2000;**1459**:266–73. https://doi.org/10.1016/S0005-2728(00)00161-4

93. Cross R, Lloyd D, Poole R *et al*. Enzymatic removal of nitric oxide catalyzed by cytochrome *c′* in *Rhodobacter capsulatus*. *J Bacteriol* 2001;**183**:3050–4. https://doi.org/10.1128/JB.183.10.3050-3054.2001

94. Caranto JD, Lancaster KM. Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proc Natl Acad Sci* 2017;**114**:8217–22. https://doi.org/10.1073/pnas.1704504114