Distribution of alternative routes for completing denitrification in microbial genomes and metagenomes of global biomes

Grace Pold[1#], Aurélien Saghaï[1], Christopher M Jones[1], and Sara Hallin[1*]

[1]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala

* Corresponding author: sara.hallin@slu.se

Current address: Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala

Running title: Global ecology of denitrification trait

Keywords: comparative genomics; comparative metagenomics; community ecology; division of labor; global biomes; nitrous oxide; split pathtways

Abstract

Diverse microorganisms can execute one or more steps in denitrification, during which nitrate or nitrite is successively reduced into nitric oxide, nitrous oxide, and ultimately dinitrogen. Many of the best-characterized denitrifiers are “complete” denitrifiers capable of executing all steps in the process, but whether they dominate in natural communities and what environmental factors drive the distribution of complete vs. partial denitrifiers remain unclear and incomplete in a global context. To address this, we conducted a comparative analysis of denitrification genes in 48,284 genomes and assessed their environmental distribution in 3,991 metagenomes covering all major biomes. We determined that the genomic potential to initiate denitrification is more common than the potential to terminate it, particularly in high nitrate or nitrite environments, and that partial denitrifiers outnumber complete denitrifiers across the majority of biomes. Our results further indicate that complete denitrifiers are more likely to be fast-growing organisms compared to partial denitrifiers, indicating the former may grow opportunistically when conditions are favorable. Together, our results indicate an environmental footprint on the gain and loss of denitrification genes which favors the genomic potential for partial over complete denitrification in most biomes, in line with the scarcity of high carbon input rates that might favor organisms encoding complete denitrification.

Main

The potent greenhouse gas and ozone depleting agent nitrous oxide (N_2O) is produced in several microbial N-cycle processes, of which denitrification plays a dual role in both producing and consuming N_2O. In this facultative anaerobic microbial respiratory pathway, nitrate (NO_3^-) or nitrite (NO_2^-) are used as electron acceptors and successively reduced into nitric oxide (NO), N_2O, and dinitrogen (N_2) by a diverse range of predominantly bacterial species. As the genes encoding the enzymes involved in these reductive steps can be independently gained and lost [1–3], they are found in different combinations among genomes [4] such that denitrification can be executed by complete denitrifiers with genes encoding enzymes for all steps or, alternatively, split between multiple partial denitrifier community members. Recent studies suggest that genetically complete denitrifiers are scarce among genomes assembled from environmental data [5–10]. However, much of our understanding of the physiology and regulation of denitrification is based on studies of a very phylogenetically restricted set of complete denitrifiers [11–16]. Thus, there is pressing need for a broader look at the prevalence, diversity and ecology of organisms genetically capable of complete and partial denitrification, particularly considering their differential roles in serving as sources and sinks of N_2O [17].

Split pathways divide the protein production costs for a pathway between multiple cells and may enable higher ATP flux [18,19], increasing the specific growth rate. Although this could be a fitness advantage under nitrite-rich conditions, ATP yields per molecule nitrate or nitrite would be lower for partial compared to complete denitrification. This is expected to lead to partial denitrification being favored when electron acceptor supply is not limiting, but complete denitrification being favored when the flux of electron donors such as carbon

compounds is high relative to electron acceptor availability. Splitting the denitrification pathway also increases dependencies between community members, not only because the denitrification product of one organism serves as the electron acceptor for another [20], but also because the regulation of specific reactions may rely on intermediates that those executing the reaction themselves cannot use [21]. Further, there is less buildup of toxic intermediate compounds, which may for instance allow microorganisms only capable of carrying out the initial steps of denitrification avoid the cytotoxic effects of NO_2^- under low pH environments [11,22]. This jointly indicates that split pathways may open more niches, thereby supporting a greater diversity of denitrifiers compared to complete denitrification [14].

Here we combined comparative genomics with a global survey of denitrification gene ratios in environmental metagenomes to evaluate the diversity of denitrifier types and their distribution across engineered, environmental and host-associated biomes. We posited that environmental factors favoring or disfavoring execution of the different steps leave a genetic footprint in the form of presence and absence of genes encoding the enzymes responsible for them. We used a framework wherein partial denitrifiers includes genetically defined "initiators" encoding genes involved in NO_2^- but not N_2O reduction, and "terminators" that encode genes for N_2O but not NO_2^- reduction. Nitrate reductases, catalyzing the first step in the denitrification pathway, were excluded from the analyses because they are found in a wide variety of microorganisms and their activity often occurs in isolation from the characteristic gas-producing denitrification process itself [23–25]. Instead, we used NO_2^- reduction, the decisive step of denitrification [23], as the first point of our analysis. We evaluated 48,284 bacterial and archaeal genomes, metagenome assembled genomes (MAGs), and single cell genomes (SCGs) and 3,991 metagenomes derived from all major biomes to

determine the frequency of a complete vs. partial denitrification pathway and the genetic potential to initiate vs. terminate denitrification at the genome and community level. We interpret our results in the context of resource availability and the genomic traits associated with the denitrifier types in our framework.

Results

Partial denitrification dominates microbial genomes

To assess the prevalence of the complete denitrification pathway among bacterial and archaeal genomes, we compiled a database of 18,867 isolate (>80% complete, <5% contaminated), 28,863 MAGs, and 509 SCGs derived from studies in engineered, environmental, and host-associated biomes (Supplementary Data 1; Fig. 1a). Genes for the denitrification enzymes NirK (*nirK*) and NirS (*nirS*) catalyzing NO_2^- reduction to NO, all proposed NO-reducing heme copper oxidases (i.e. Nor (*nor*) [25]), and NosZ (*nosZ*I, *nosZ*II) catalyzing N_2O reduction were annotated using hidden Markov models trained on manually curated databases of these genes [26].

Twenty three percent of the remaining genomes carrying the genes *nir* and/or *nosZ* (i.e. "denitrifiers") encoded genes for complete NO_2^- reduction to N_2 ("complete denitrifiers"; Fig. 1b), with the complete denitrification trait twice as prevalent among isolates compared to MAGs (30% vs. 16%, respectively, Supplementary Data 2). However, there was considerable variation between phyla (Supplementary Fig. 1). The complete denitrification trait was found in 26 different bacterial and 2 archaeal phyla (Fig. 1c, Supplementary Data 3), but among phyla represented by at least 100 denitrifier genomes it was most prevalent within Pseudomonadota (38% of

denitrifiers), Campylobacterota (36%) and Myxococcota (31%, Fig. 1d). On the other hand, partial denitrifiers could be found in 68 bacterial and 6 archaeal phyla, and were most prevalent within Actinobacteriota (100%) and Verrucomicrobiota (97%) (Supplementary Data 3).

Using our framework of initiators and terminators, we found that initiators encoding *nir* but not *nosZ* were broadly more prevalent than terminators having *nosZ* but not *nir* (46% vs. 26% of genomes with denitrification genes, respectively), which may represent the potentially divergent functions for NO_2^- versus N_2O reduction as a means of growth or detoxification vs. maintenance [27]. Actinobacteriota (98% initiators, 1% terminators) were overrepresented in the first group, while Campylobacterota (6%, 58%) and Bacteroidota (16%, 61%) had a higher proportion of terminators than initiators (Fig. 1d; Supplementary Data 3). We also found genomes carrying *nir* and *nosZ* but not *nor*, suggesting a third category: the "initiator-terminators". They accounted for just 4% of genomes with denitrification genes, including 25% of Gemmatimonadetes, 10% of Verrucomicrobiota, 11% of Chloroflexota, and 9% of Bacteroidota. It is possible that these organisms either depend on NO reduction by other organisms or produce N_2O via pathways that do not depend on Nor [28–31].

Among all genomes carrying *nosZ* (i.e. terminators, initiator-terminators, and complete denitrifiers), those encoding clade I *nosZ* were more likely to encode the complete denitrification trait compared to genomes encoding clade II (69% in clade I vs. 25% in clade II; Supplementary Fig. 2), as previously observed [4]. Within clade II, complete denitrification was most prevalent in Pseudomonadota and Aquificota, with 76% of

Pseudomonadota and 88% of Aquificota encoding complete denitrification compared to an average of 19% in the remaining phyla. This suggests that the difference in clade I vs. clade II prevalence could be a proxy for complete vs. partial denitrifier prevalence within the N_2O reducing community of environmental samples, but not if clade II is dominated by Pseudomonadota or Aquificota.

Inferred ecological preferences of complete vs. partial denitrifiers

We subsequently assessed whether the genomic potential for complete denitrification or for initiating rather than terminating denitrification was associated with predicted growth rate and resource availability. Using the codon bias-based maximum growth rate prediction tool gRodon [32], we categorized genomes into fast- (growth rate faster than 0.2 h^{-1}) and slow-growing (growth rate lower than 0.2 h^{-1}) [32]. Rapid growth was more common among complete denitrifiers (78%) compared to initiators (68%), terminators (61%) and initiator-terminators (53%; Fig. 2a) and for genomes encoding clade I (77%) compared to clade II NosZ (57%; Fig. 2b). Similarly, estimated median growth rates of complete denitrifiers (0.36 h^{-1}) were greater than that of initiators (0.31 h^{-1}), terminators (0.24 h^{-1}) and initiator-terminators (0.21 h^{-1}; overall Kruskall-Wallis χ^2 = 401.2, Dunn $P < 0.001$ in all cases).

We could make a more direct assessment of the degree to which complete and partial denitrifiers were associated with high or low resource availability by using CoverM [33] to map reads from the TARA Oceans [34] project to MAGs derived from the same data [35]. No subset of the metagenome data we collected or other databases we screened had a sufficient number of MAGs and environmental data to allow for this analysis. Nitrate

was negatively correlated with dissolved oxygen availability (Spearman $\rho = 0.53$, $P < 2.2 \times 10^{-16}$) and was used as a proxy for denitrification electron acceptor availability. Chlorophyll concentration was used as a proxy for electron donor supply because of its positive association with gross primary productivity [36]. The proportion of reads mapping to the collection of 170 MAGs in the dataset differed by sample and ranged from 5 to 30%. By correlating the proportion of reads mapping the denitrifier types among the MAGs (7 complete, 13 initiators, 14 terminators, 134 non-denitrifiers and 2 initiator-terminators) with chlorophyll and NO_3^- concentrations in different samples and standardizing this to the maximum observed relative abundance of each MAG, we could calculate a standardized environmental response. We found that terminators and complete denitrifiers tended to have a greater increase in relative abundance with increasing NO_3^- than non-denitrifiers, but there was no difference between complete denitrifiers and partial denitrifiers (Fig. 3a). Mean standardized chlorophyll and chlorophyll: NO_3^- ratio responses were similar across all groups of organisms, independent of denitrifier type (Fig. 3b,c). One possible reason for the apparent lack of correlation between denitrifier type and resource availability is that these bulk measurements are not representative of the microhabitat that denitrifiers experience. Further, the genomes in our dataset may not represent the most prevalent denitrifiers or sites in the oceans as denitrification activity is concentrated on particulate organic matter[37], which was depleted by pre-filtering in these samples. Nonetheless, the results from this pelagic marine dataset do not indicate that resource availability affects the prevalence of complete or partial denitrifier genomes, but shows that denitrifiers overall are favored by higher NO_3^- concentrations compared to non-denitrifiers.

Prevalence of denitrification initiators and terminators varies across global biomes

Next, we assessed variation in denitrification initiator and terminator prevalence and complete and partial denitrifiers among terminators (*i.e.* N_2O reducers) at the community level in a broad range of environments. We screened 3,991 metagenomes derived from soil, aquatic, engineered and host-associated biomes for *nirK*, *nirS*, and *nosZ* clades I and II (Supplementary Data 4) using an HMM-based search and phylogenetic placement approach [40]. The difference in the copy number of *nosZ* and *nir* genes normalized to Gbp sequenced in each metagenome (*δnos-nir*) was used as a proxy for the genetic potential for termination compared to initiation of denitrification, and *nosZ* clade I vs. clade II (*δnosZ*I-*nosZ*II) as a proxy for complete denitrifier dominance within the N_2O reducing community. Delta values were chosen for their ability to account for differences in overall denitrifier prevalence between biomes, including the case where one or both of the denitrification genes being compared were not detected.

Across the majority of biomes, *nir* gene fragments were typically more abundant than *nosZ* (Fig. 4a), indicating a greater or similar potential to initiate than terminate denitrification within the denitrifier communities. Marine mats served as an exception, and *nosZ* prevalence exceeded that of *nir* by 60%. This could reflect higher rates of NO_3^- assimilation or low inputs of NO_2^- from nitrification [38], as suggested by the overall low prevalence of *nir* and lack of nitrifier-denitrifier associated *nirK* clades in these samples [26]. *nosZ* Clade II was more prevalent than *nosZ* Clade I in all biomes considered, except for sewage (*δnosZ*I-*nosZ*II < 0; Fig. 4b). This indicates that partial denitrifiers generally dominate terminator communities because partial denitrification is more prevalent among Clade II *nosZ*. This indication is further supported by our observation that complete denitrifier poor phyla, such as

Bacteroidetes, Chloroflexota and Gemmatimonadota, dominated *nosZ* clade II across biomes (Fig. 4c,d). Therefore, the majority of organisms capable of terminating denitrification in global biomes do not also initiate it. This is likely to be particularly the case in croplands, marshes, and activated sludge from municipal wastewater treatment plants, the biomes where *δnosZ*I-*nosZ*II is lowest for terrestrial, marine, and engineered environments (Fig. 4b). Moreover, the dominance of *nosZ*II coincides with the highest total *nosZ* and *nir* gene abundances (Supplementary Data 5), indicating conditions overall more favorable for denitrification promote N_2O reducers that are partial rather than complete denitrifiers. A notable exception to this pattern is sewage communities, which were dominated by clade I *nosZ* despite having high overall denitrification gene fragment prevalence. We also assessed the correlation between the diversity of denitrifiers based on balance-weighted phylogenetic diversity of the dominant nitrite reductase, *nirK* and potential initiator-terminator or clade I:II differential abundance 26. We found that *nirK* phylogenetic diversity was positively correlated with *δnos-nir* in soils (Spearman's $\rho = 0.46$, $P < 2.2 \times 10^{-16}$) and marine samples ($\rho = 0.19$, $P = 1.18 \times 10^{-11}$), while diversity was positively correlated with *δnosZ*I-*nosZ*II in soils ($\rho = 0.29$, $P < 2.2 \times 10^{-16}$) and uncorrelated in marine samples ($\rho = 0.03$, $P = 0.35$). These results indicate that more diverse denitrifier communities occur where complete denitrifiers and terminators are relatively more prevalent compared to initiators, such that partial denitrification is not intrinsically associated with greater niche partitioning.

Environmental drivers of *nir* vs. *nosZ and nosZ* clade prevalence were assessed using random forest modeling on the largest soil and marine datasets with complete metadata represented in our analysis [44,45]. We generated accumulated local effect plots to show the main effect of the target variable on predicted *nos-nir* (Fig. 5) and *δnosZ*I-*nosZ*II (Fig. 6), while accounting for

the other predictors. Models explained less than half of the variance in gene ratios, except for *δnosZ*I-*nosZ*II in soils. Soil organic carbon content up to ~1% positively affected δ*nos-nir*, while the effect of soil NO_3^- content was always negative (Fig. 5a). *nosZ* was also predicted to become less prevalent compared to *nir* as available phosphorus increased, indicating that high nutrient conditions in soils favor NO_2^- reducers over N_2O reducers. Within the marine dataset, predicted δ*nos-nir* was lower at the highest concentrations of NO_3^- and NO_2^- compared to the lowest concentrations, and showed a dramatic shift in prediction from a decreasing to increasing relationship with temperature at 22 ^oC (Fig. 5b).

Soil pH, available phosphorus, potassium, clay content and NO_3^- all had a negative effect on predicted δ*nosZ*I-*nosZ*II, though only pH had a negative effect over the entirety of measured values (Fig. 6a). Increases in elevation at low altitude were associated with increased predicted *δnosZ*I-*nosZ*II, but additional increases in elevation were not associated with further increases. Soil organic carbon content did not affect predicted *δnosZ*I-*nosZ*II. In the marine study, temperature and ammonium concentrations had opposing unimodal effects on predicted values over their observed ranges (Fig. 6b).

Discussion

By combining a comparative analysis of isolate, single cell and metagenome-assembled genomes with a broad metagenomic survey, we provide unprecedented insight into the prevalence of the complete and partial denitrification trait among global microbial communities. In contrast to a previous comparative genomics study based on a small set of isolated microbes [4] but consistent with more recent studies using MAGs [5,9,25,39,40], we found

that the genetic potential for complete denitrification is less common than that for partial denitrification, among both genome-sequenced microorganisms and environmental communities from all major biomes.

Various explanations exist for the dominance of partial denitrifiers. Our results did not support the hypothesis that partial denitrification enables greater overall niche partitioning. Rather, complete denitrifiers may be adapted to flexibly take advantage of resources varying in time and space, while partial denitrifiers maintain slower but more stable growth rates. This is supported by our observation that complete denitrifiers are more likely to have fast maximum predicted growth rates compared to partial denitrifiers. Under conditions supporting rapid growth, complete denitrifiers may benefit from using available NO_3^-/NO_2^- more efficiently, while under rapidly fluctuating conditions they have the flexibility to use whichever denitrification terminal electron acceptor is most readily available. Fast-growing microorganisms are relatively rare in natural ecosystems [32,41,42] but overrepresented among isolates due to their ability to grow readily under standard resource-rich lab conditions. This may explain both the over-representation of complete denitrifiers among model denitrifiers and similar prevalence of complete denitrification among isolates observed here as in a previous isolate-centric analysis [4]. By contrast, MAGs lack cultivation bias and better represent the diversity of organisms found in the environment, but are susceptible to partial and incorrect assembly, particularly when it comes to accessory genes (including denitrification genes [1,43]) held on plasmids and genomic islands [44]. Nonetheless, the prevalence of the partial denitrifier dominated clade II *nosZ* genes and imbalance

between *nir* and *nosZ* across the majority of global biomes, support our inference that partial denitrifiers are more prevalent than complete denitrifiers.

Partial denitrifiers may also dominate communities due to the presence of a rate-efficiency tradeoff. Energy flux is slower through long pathways but potentially enables more complete usage of the terminal electron acceptor when it is limiting [19]. Accordingly, complete denitrification is expected to decline as NO_3^- increases or as C:NO_3^- and C:NO_2^- ratios decrease. Our random forest modeling inferred that the N_2O reducing community in higher NO_3^-/NO_2^- soils became more dominated by partial denitrifiers, i.e. those carrying *nosZ* clade II. Further, the denitrifying community shifted in favor of NO_2^- reduction over N_2O reduction as NO_3^- levels increased in both soils and marine waters. This is consistent with the predicted prevalence of the competing pathway dissimilatory nitrate reduction to ammonium (DNRA) compared to denitrification in this same dataset [45], and indicates that low availability of electron acceptor or a high electron donor:acceptor ratio favors more complete community-level reduction of NO_2^- via either denitrification or DNRA. Nonetheless, this pattern did not hold in the MAG read mapping analyses completed for marine samples. One reason could be a greater use of alternative non-nitrogenous terminal electron acceptors, considering the greater reductase gene diversity often found in organisms encoding the marine-enriched *nirS* compared to the soils-enriched *nirK* nitrite reductase variant [26]. Moreover, the favorability of having, expressing, and using enzymes for the complete denitrification pathway also depends on a whole host of factors not readily captured by bulk-level metrics of resource availability such as NO_3^- and chlorophyll levels [46,47]. This is evident in the generally small proportions of variance explained by our random

forest models which used dominant quantitative drivers of denitrification as predictors. Targeting the chemical composition of organic matter in addition to its quantity thus represents a promising avenue to further explore to gain understanding into denitrifier ecology.

Initiation of the denitrification pathway at the step of nitrite reduction was twice as prevalent as the ability to terminate it. This aligns with previous genome and site-specific studies [5,9,39,40]. Denitrification does not occur in isolation, and its steps may occur in response to its intermediates being produced by other pathways, such as NO from host immune response or microbial detoxification of NO_2^-, and N_2O from nitrification, ammonification [30,31], detoxification of NO [48,49], or even biotically-mediated chemodenitrification [50]. While NO_2^- also serves as a substrate for NO_2^- oxidation, anammox and dissimilatory nitrite reduction to ammonium, NosZ is the only enzyme known to reduce N_2O. It may be that the extent of N_2O reduction capacity is larger than we know, as indicated by the rare (0.4% of *nosZ* here) recently proposed [10,51] clade III NosZ which forms the outgroup in our phylogeny. Alternatively, it can be that nitrite reductases are involved in processes beyond denitrification [52,53] or that inter-pathway competition for NO_2^- may favor the presence of initiators ready to consume NO_2^- in denitrification rather than letting it enter other pathways. The latter may be the case in heavily fertilized soils where both initiation and termination specialists prevail, as evidenced by both lower δ*nir*-*nos* and δ*nosZ*I-*nosZ*II under these conditions. In biomes such as activated sludge where N_2O is produced by nitrifier-denitrification and chemodenitrification [54,55], terminators may be particularly prevalent within the N_2O reducing community. Hence, our results indicate that the prevalence of denitrification

genotypes within and between environments should be considered in light of not only the potential organism level advantages to having a particular denitrification genotype under given abiotic conditions, but also how that organism may interact with denitrifiers and non-denitrifiers alike producing and consuming NO_2^-, NO and N_2O. Future work addressing phenotypic plasticity of denitrifiers under realistic environmental conditions are necessary to resolve this uncertainty and untangle drivers of denitrifier community assembly.

Methods

Denitrification enzyme databases

A hidden Markov model for NosZ was generated by updating an alignment from Graf et al. [4] with NosZ sequences present in RefSeq release 95, following the procedure described in Pold et al. [26]. After manually checking the alignment for conserved residues [56] and trimming the alignment to its conserved core, we built a phylogeny using IQ-TREE with the best model identified by modelfinder (LG+R10). Alignments and phylogenies for NirK and NirS were obtained from Pold et al. [26]. The Nor reference database was derived from Murali et al. [57] and consisted of 67 Nor sequences (cNOR, qNOR, bNOR, eNOR, gNOR and sNOR, and nNOR) and 865 other heme copper oxidases.

Genome database generation and annotation

We compiled genomes from large databases consisting primarily of MAGs derived from host-associated, environmental, and engineered biomes, including human and other animal microbiomes, soil, aquifer, marine, freshwater, plant-associated, food,

sewage, subway and wastewater treatment plants [9,58–60,35,61–66]. Unless genomes had already been dereplicated by the authors [35,59], we deemed that genomes sharing > 96% ANI using FastANI v.1.2 [67] belonged to the same OTU and the assembly with the highest quality score (estimated using CheckM v. 1.0.11) was kept. All assemblies with lower than 80% completeness or >5% contamination were discarded, leaving 48,284 genomes for further analysis. Taxonomy of genomes lacking such data in their associated metadata was assigned using GTDB/TK (v. 1.5.0; [68]) against GTDB v. 202.

We identified genes for NirK, NirS, Nor, and NosZ in each genome in our database using an HMMsearch [69] against the corresponding amino acid databases as described above. After aligning potential proteins to the reference, we imported them into ARB v.7.0 [70] and used a combination of phylogeny building with FastTreeMP v.2.1 [71] and manual curation to exclude reads lacking conserved ligand binding domains. Nor and NosZ were also classified according to protein class (i.e. bNOR, cNOR, eNOR, gNOR, nNOR, qNOR and sNOR for Nor and halophiles, clade I and clade II for NosZ). We then classified each genome into non-denitrifiers (lacking *nir* and *nosZ*), complete denitrifiers (*nir*, *nor*, and *nos*), and partial denitrifiers (lacking *nir* or *nosZ*), with partial denitrifiers further divided into "initiators" (*nir* +/- *nor*), terminators (*nosZ* +/- *nor*) and initiator-terminators (*nir* and *nosZ)*. Genomes from anammox (n=25) and archaeal nitrifier genera (n=103) carrying *nir* were classified as non-denitrifiers because their nitrite reductase is not thought to be involved in denitrification [72–74].

Unless otherwise noted, all subsequent analyses were completed in R v.4.1 [75]. The maximum growth rate of each taxon was estimated using gRodon [32], which used codon

usage bias of ribosomal proteins annotated in the genomes using prokka [76]. Differences in growth rates between denitrifier types were assessed using Kruskall-Wallis tests followed by Dunn tests with Benjamini-Hochberg correction for multiple testing, where all maximum growth rates less than 0.2 doublings h^{-1} were given the same rank. This cutoff was needed because codon use bias saturates with growth rates slower than this [32]. We elected not to exclude the effect of shared organism ancestry on growth rate because we were interested in whether other genomic traits are associated with the complete denitrification trait. Whether those traits coexist due to being present in an ancestor or acquired in coordination with the complete denitrification trait, and whether those traits coexist due to being present in an ancestor or acquired in coordination with complete denitrification is irrelevant.

Metagenome and metadata collection

To assess the dominance of complete vs. partial denitrifier communities across biomes, we selected metagenomes with at least 100,000 reads ≥150 nt, primarily from larger sampling campaigns where uniform metadata were available. The majority of soil metagenomes came from the Australian Microbiome Project/BASE [77], the National Ecological Observatory Network [78], Topsoils Microbiome Project [79], Long-Term Soil Productivity experiment [80], and the Stordalen Mire [81]. The majority of aquatic metagenomes came from the Australian Microbiome Project [82], Linnaeus Microbial Observatory [83], the Amazon Continuum Project [84] and BATS, GEOTRACE and HOTS [85]. Engineered metagenomes included those from drinking water [86] sewage [87], and wastewater treatment plants [60]. The host-associated metagenomes dataset consists primarily of plant-associated sequences, such as from leafy greens [88], beans [89], citrus

rhizosphere [90], switchgrass [91] and *Arabidopsis* [92], though we also included a handful of sequences from invertebrates such as sponges and gutless marine worms. For unpublished metagenomes where contact details were provided, we emailed authors to request permission to use them. Where multiple sequencing runs or files existed for the same sample, we combined the fastq files prior to searching. We used only the forward read in the case of paired end reads, and only the first 150nt of any reads longer than that. In total, our dataset comprised 3,991 metagenomes, with 1,489 from aquatic habitats, 1,642 from soil, 658 host-associated, and 202 from engineered habitats (Supplementary Data 4).

Biomes were assigned based on The Nature Conservancy Terrestrial Ecoregions for terrestrial samples [93] and latitude for marine samples (polar: latitude > 60°; westerlies: latitude 30-60°; trades: latitude 0-30°). Terrestrial biome assignment required use of the packages sp v.1.4-5 [94], rgdal v.1.523 [95], and rgeos v0.5-5 [96]. Soils under cultivation were excluded from the biome-based approach and instead categorized as croplands.

Dominance of complete vs. partial denitrifier communities based on proxy gene searches across biomes

We used GraftM [97] to identify denitrification genes (*nirK*, *nirS* and *nosZ*) in the metagenomes, as previously described [26]. Briefly, GraftM uses a two-step process in which a HMM search identifies candidate reads, followed by phylogenetic placement on a reference tree using pplacer (v.1.1.19; [98]). We used the accumulate function in gappa[99] to find the position on the tree where at least 95% of the mass for each read descended from, and excluded all reads with any mass placed in the outgroup. We

further excluded sequences placed in the non-denitrifying anammox NirS clade 1h, archaeal nitrifier NirK clades 2 and 4, as well as the eukaryotic NirK clade 1e [22]. This metagenome search depended on: hmmer (v.3.2.1; [69]), OrfM (v.0.7.1; [100]), bbtools (v.38.90; "https://sourceforge.net/projects/bbmap/"), lbzip2 (v.2.5; "https://lbzip2.org/"), and fxtract (v.2.3; "https://github.com/ctSkennerton/fxtract").

We validated these search and placement and post-processing parameters by generating 150 nt fragments of full-length sequences that were picked up by the NirK, NirS and NosZ HMMs, which in addition to the target proteins also respectively included homologous multicopper oxidase proteins from Cyanobacteriota and Thermoprotetoa, NirN and NirF, and various yet to be characterized NosZ-like proteins. Sensitivity is the fraction of total ingroup fragments searched which were placed in the ingroup, and specificity is one minus the fraction of outgroup fragments incorrectly placed in the ingroup (or, in the case of *nosZ*, in a clade other than the target clade). Sensitivity was 76% for *nirK*, 93% for *nirS*, 91% for clade I *nosZ* and 94% for clade II *nosZ*. Specificity was 97% for *nirK*, 100% for *nirS*, 100% for clade I *nosZ* and 100% for clade II *nosZ*.

We compared gene prevalence across biomes using the absolute difference in counts of clade I and II *nosZ* or nir and *nosZ* genes standardized by the number of Gbp sequenced to account for differences in sequencing depth. For *nir* vs. *nosZ* abundance we used Eq. 1 such that positive values denote *nosZ* is more abundant than *nir*:

$$\delta nos - nir = 10^9 \times (nosZI + nosZII - nirS + nirK)/(150 \times \text{reads}) \qquad (1)$$

For *nosZ,* we used Eq. 2 such that positive values denote clade I more common than clade II:

$$\delta nosZI\text{-}nosZII = 10^9 \times (nosZI - nosZII)/(150 \times \text{reads}) \quad (2)$$

This method of calculating gene prevalence enables the abundance of both genes to be considered concurrently. This is particularly relevant for the 14% of metagenomes where *nosZ*I was not detected but *nosZ*II was found in 1-13,041 copies, and the 0.8% of metagenomes where *nosZ* was not found but 3-28 copies of *nir* were identified.

Environmental correlates with biome gene counts and denitrifier MAGs

We used the TARA Oceans [34] MAGs present in the OceanDNA MAG database to test the hypothesis that complete denitrifiers are more abundant under conditions with high electron donor to acceptor ratios. These MAGs were chosen because the database is sufficiently large that multiple MAGs from each denitrifier type category were present, and where multiple conspecific MAGs were present, those with inconsistent denitrification gene repertoires could be excluded. After quality-controlling the raw paired-end reads with cutadapt (-m 100 -q 20 –max-n 0 –trim-n), we mapped them to all MAGs using CoverM (v. 0.6.1; [33]; -m relative abundance –min-read-aligned-percent 0.75 –min-read-percent-identity 0.95 –min-covered-fraction 0) then determined the total relative abundance of each MAG-OTU in each sample (95% ANI). Only the reads for the 3 µm filter fraction were used in this analysis to allow for bigger cells or cell aggregates than using the 0.22 µm fraction. We subsequently determined the standardized relative abundance by dividing the observed relative abundance within each sample to the maximum relative abundance observed across all samples. We then fit a linear model predicting the standardized relative abundance of each MAG OTU based on the combination of NO_3^- and chlorophyll concentrations (i.e. NO_3^- after accounting for chlorophyll, chlorophyll after accounting for NO_3^-, and the logarithm of

chlorophyll: NO_3^-). The mean slope and standard error for each MAG OTU were extracted and used to calculate the weighted mean response and 95% confidence intervals for each denitrifier type. Differences in mean response to NO_3^-, chlorophyll and their ratios between denitrifier types were established based on non-overlapping 95% confidence intervals.

We evaluated the relationship between differences in normalized counts and resource availability using random forests. This analysis was restricted to metagenomes from the BASE and Australian marine projects[77,85] as examples of studies with a large number of metagenomes with uniformly-collected metadata for variables associated with denitrification, including NO_3^-/NO_2^- levels, pH, carbon, and micronutrients. A pre-selection of variables for inclusion in the model was completed to exclude collinear variables (Spearman's $\rho > 0.7$ or variance inflation factor > 4); in the collinear groups, the variable hypothesized to be most strongly and directly explicable of gene ratio was retained (Supplementary Fig. 3; Supplementary Tables 1 and 2). We subsequently used VSURF v. 1.1.0[101] to identify the best predictors for each ratio and biome, and the randomForest package v. 4.7.1-1[102] to model the relationship between them. A grid search was used to find the combination of tuning parameters yielding lowest out-of-bag root-mean-square error. Accumulated local effects plots were generated to visualize results using iml v. 0.11 [103].

Data availability

All genomes and metagenomes used in this paper are publicly available in NCBI, IMG, or other sources under the identifiers denoted in Supplementary Data 1 and 4. Hidden Markov

Models and reference databases used for denitrification gene searches are available in FigShare (10.6084/m9.figshare.23913078 for NirK and NirS and 10.6084/m9.figshare.24922776 for NosZ).

References

1. Alvarez, L., Bricio, C., Gómez, M. J. & Berenguer, J. Lateral transfer of the denitrification pathway genes among *Thermus thermophilus* strains. *Appl. Environ. Microbiol.* **77**, 1352–1358 (2011).
2. Rees, E., Siddiqui, R. A., Köster, F., Schneider, B. & Friedrich, B. Structural gene (*nirS*) for the cytochrome cd1 nitrite reductase of *Alcaligenes eutrophus* H16. *Appl. Environ. Microbiol.* **63**, 800–802 (1997).
3. Schwintner, C., Sabaty, M., Berna, B., Cahors, S. & Richaud, P. Plasmid content and localization of the genes encoding the denitrification enzymes in two strains of *Rhodobacter sphaeroides*. *FEMS Microbiol. Lett.* **165**, 313–321 (1998).
4. Graf, D. R. H., Jones, C. M. & Hallin, S. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. *PLoS ONE* **9**, e114118 (2014).
5. Diamond, S. *et al.* Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat. Microbiol.* **4**, 1356–1367 (2019).
6. Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E. & Konstantinidis, K. T. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Appl. Environ. Microbiol.* **84**, e01646-17 (2018).
7. Dove, N. C., Torn, M. S., Hart, S. C. & Taş, N. Metabolic capabilities mute positive response to direct and indirect impacts of warming throughout the soil profile. *Nat. Commun.* **12**, 1–13 (2021).

8. Karthikeyan, S. *et al.* Metagenomic characterization of soil microbial communities in the Luquillo experimental forest (Puerto Rico) and implications for nitrogen cycling. *Appl. Environ. Microbiol.* **87**, e00546-21 (2021).
9. Pessi, I. S. *et al.* In-depth characterization of denitrifier communities across different soil ecosystems in the tundra. *Environ. Microbiome* **17**, 30 (2022).
10. Mosley, O. E. *et al.* Nitrogen cycling and microbial cooperation in the terrestrial subsurface. *ISME J.* **16**, 2561–2573 (2022).
11. Lilja, E. E. & Johnson, D. R. Segregating metabolic processes into different microbial cells accelerates the consumption of inhibitory substrates. *ISME J.* **10**, 1568–1578 (2016).
12. Lilja, E. E. & Johnson, D. R. Substrate cross-feeding affects the speed and trajectory of molecular evolution within a synthetic microbial assemblage. *BMC Evol. Biol.* **19**, 1–11 (2019).
13. Dolinšek, J., Ramoneda, J. & Johnson, D. R. Initial community composition determines the long-term dynamics of a microbial cross-feeding interaction by modulating niche availability. *ISME Commun.* **2**, 1–10 (2022).
14. Takahashi, K. *et al.* Denitrification in low oxic environments increases the accumulation of nitrogen oxide intermediates and modulates the evolutionary potential of microbial populations. *Environ. Microbiol. Rep.* doi:https://doi.org/10.1111/1758-2229.13221.
15. Vitale, A. *et al.* Mapping of the denitrification pathway in *Burkholderia thailandensis* by genome-wide mutant profiling. *J. Bacteriol.* **202**, e00304-20 (2020).
16. Zumft, W. G. & Kroneck, P. M. H. Respiratory Transformation of Nitrous Oxide (N_2O) to dinitrogen by Bacteria and Archaea. In *Advances in Microbial Physiology* (ed. Poole, R. K.) vol. 52 107–227 (Academic Press, 2006).
17. Jones, C. M. *et al.* Recently identified microbial guild mediates soil N_2O sink capacity. *Nat. Clim. Change* **4**, 801–805 (2014).

18. Heinrich, R., Schuster, S. & Holzhütter, H.-G. Mathematical analysis of enzymic reaction systems using optimization principles. *EJB Rev. 1991* 167–187 (1991).
19. Kreft, J.-U., Griffin, B. M. & González-Cabaleiro, R. Evolutionary causes and consequences of metabolic division of labour: why anaerobes do and aerobes don't. *Curr. Opin. Biotechnol*. **62**, 80–87 (2020).
20. Wang, Y., Gao, H. & F. Wells, G. Integrated omics analyses reveal differential gene expression and potential for cooperation between denitrifying polyphosphate and glycogen accumulating organisms. *Environ. Microbiol.* **23**, 3274–3293 (2021).
21. LaSarre, B., Morlen, R., Neumann, G. C., Harwood, C. S. & McKinlay, J. B. Nitrous oxide reduction by two partial denitrifying bacteria requires denitrification intermediates that cannot be respired. 2022.06.21.497020 Preprint at https://doi.org/10.1101/2022.06.21.497020 (2023).
22. Crocker, K. *et al.* Environmentally dependent interactions shape patterns in gene content across natural microbiomes. *Nat. Microbiol.* **9**, 2022–2037 (2024).
23. Mahne, I. & Tiedje, J. M. Criteria and methodology for identifying respiratory denitrifiers. *Appl. Environ. Microbiol.* **61**, 1110–1115 (1995).
24. Sun, Y., De Vos, P. & Heylen, K. Nitrous oxide emission by the non-denitrifying, nitrate ammonifier *Bacillus licheniformis*. *BMC Genomics* **17**, 68 (2016).
25. Zhang, I. H. *et al.* Genome-resolved metagenomics reveals abundant nitrate reducers and partitioning of nitrite usage within global oxygen deficient zones. 2023.03.02.530666 Preprint at https://doi.org/10.1101/2023.03.02.530666 (2023).
26. Pold, G. *et al.* Phylogenetics and environmental distribution of nitric oxide forming nitrite reductases reveals their distinct functional and ecological roles. *ISME Commun.* **4**, ycae020 (2024).
27. Park, D., Kim, H. & Yoon, S. Nitrous oxide reduction by an obligate aerobic bacterium, *Gemmatimonas aurantiaca* Strain T-27. *Appl. Environ. Microbiol*. **83** e00502-17 (2017).

28. Costa, C. *et al.* Regulation of the hexaheme nitrite/nitric oxide reductase of *Desulfovibrio desulfuricans*, *Wolinella succinogenes* and *Escherichia coli*. *FEBS Lett.* **276**, 67–70 (1990).
29. Kim, S. O., Orii, Y., Lloyd, D., Hughes, M. N. & Poole, R. K. Anoxic function for the *Escherichia coli* flavohaemoglobin (Hmp): reversible binding of nitric oxide and reduction to nitrous oxide. *FEBS Lett.* **445**, 389–394 (1999).
30. Xu, Z. *et al.* Unprecedented N_2O production by nitrate-ammonifying Geobacteraceae with distinctive N_2O isotopocule signatures. *mBio* **0**, e02540-24 (2024).
31. Simon, J. & Klotz, M. G. Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *Biochim. Biophys. Acta BBA-Bioenerg.* **1827**, 114–135 (2013).
32. Weissman, J. L., Hou, S. & Fuhrman, J. A. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci.* **118**, e2016810118 (2021).
33. Ben Woodcroft. CoverM. (2021).
34. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
35. Nishimura, Y. & Yoshizawa, S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Sci. Data* **9**, 305 (2022).
36. Morin, A., Lamoureux, W. & Busnarda, J. Empirical models predicting primary productivity from chlorophyll a and water temperature for stream periphyton and lake and ocean phytoplankton. *J. North Am. Benthol. Soc.* **18**, 299–307 (1999).
37. Wan, X. S. *et al.* Particle-associated denitrification is the primary source of N_2O in oxic coastal waters. *Nat. Commun.* **14**, 8280 (2023).

38. Coban, O., Rasigraf, O., de Jong, A. E. E., Spott, O. & Bebout, B. M. Quantifying potential N turnover rates in hypersaline microbial mats by using ^{15}N tracer techniques. *Appl. Environ. Microbiol.* **87**, e03118-20 (2021).

39. Hester, E. R., Jetten, M. S., Welte, C. U. & Lücker, S. Metabolic overlap in environmentally diverse microbial communities. *Front. Genet.* **10**, 989 (2019).

40. Valk, L. C. *et al.* Exploring the microbial influence on seasonal nitrous oxide concentration in a full-scale wastewater treatment plant using metagenome assembled genomes. *Water Res.* **219**, 118563 (2022).

41. Stone, B. W. G. *et al.* Life history strategies among soil bacteria—dichotomy for few, continuum for many. *ISME J.* **17**, 611–619 (2023).

42. Caro, T. A., McFarlin, J., Jech, S., Fierer, N. & Kopf, S. Hydrogen stable isotope probing of lipids demonstrates slow rates of microbial growth in soil. *Proc. Natl. Acad. Sci.* **120**, e2211625120 (2023).

43. Jang, J., Sakai, Y., Senoo, K. & Ishii, S. Potentially mobile denitrification genes identified in *Azospirillum* sp. Strain TSH58. *Appl. Environ. Microbiol.* **85**, e02474-18 (2019).

44. Maguire, F. *et al.* Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic islands. *Microb. Genomics* **6**, (2020).

45. Saghaï, A., Pold, G., Jones, C. M. & Hallin, S. Phyloecology of nitrate ammonifiers and their importance relative to denitrifiers in global terrestrial biomes. *Nat. Commun.* **14**, 8249 (2023).

46. Conthe, M. *et al.* Denitrification as an N_2O sink. *Water Res.* **151**, 381–387 (2019).

47. Livingston, J., Spero, M. A., Lonergan, Z. R. & Newman, D. K. Visualization of mRNA expression in *Pseudomonas aeruginosa* aggregates reveals spatial patterns of fermentative and denitrifying metabolism. *Appl. Environ. Microbiol.* **88**, e00439-22 (2022).

48. Poole, R. K. Flavohaemoglobin: the pre-eminent nitric oxide–detoxifying machine of microorganisms. *F1000Research* **9**, F1000 Faculty Rev-7 (2020).

49. Wonderen, J. H. van, Burlat, B., Richardson, D. J., Cheesman, M. R. & Butt, J. N. The nitric oxide reductase activity of cytochrome c nitrite reductase from *Escherichia coli*. *J. Biol. Chem.* **283**, 9587–9594 (2008).

50. Onley, J. R., Ahsan, S., Sanford, R. A. & Löffler, F. E. Denitrification by *Anaeromyxobacter dehalogenans*, a common soil bacterium lacking the nitrite reductase genes *nirS* and *nirK*. *Appl. Environ. Microbiol.* **84**, e01985-17 (2018).

51. He, G. *et al.* Sustained bacterial N_2O reduction at acidic pH. *Nat. Commun.* **15**, 4092 (2024).

52. Hira, D. *et al.* Anammox organism KSU-1 expresses a NirK-type copper-containing nitrite reductase instead of a NirS-type with cytochrome cd_1. *FEBS Lett.* **586**, 1658–1663 (2012).

53. Higgins, S. A., Schadt, C. W., Matheny, P. B. & Löffler, F. E. Phylogenomics reveal the dynamic evolution of fungal nitric oxide reductases and their relationship to secondary metabolism. *Genome Biol. Evol.* **10**, 2474–2489 (2018).

54. Kampschreur, M. J., Kleerebezem, R., de Vet, W. W. J. M. & van Loosdrecht, M. C. M. Reduced iron induced nitric oxide and nitrous oxide emission. *Water Res.* **45**, 5945–5952 (2011).

55. Soler-Jofra, A. *et al.* Importance of abiotic hydroxylamine conversion on nitrous oxide emissions during nitritation of reject water. *Chem. Eng. J.* **287**, 720–726 (2016).

56. Sanford, R. A. *et al.* Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc. Natl. Acad. Sci.* **109**, 19709–19714 (2012).

57. Murali, R. *et al.* Diversity and evolution of nitric oxide reduction in bacteria and archaea. *Proc. Natl. Acad. Sci.* **121**, e2316422121 (2024).

58. Buck, M. *et al.* Comprehensive dataset of shotgun metagenomes from oxygen stratified freshwater lakes and ponds. *Sci. Data* **8**, 131 (2021).

59. Nayfach, S. *et al*. A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* **39**, 499–509 (2021).

60. Singleton, C. M. *et al*. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat. Commun.* **12**, 1–13 (2021).

61. Delmont, T. O. *et al*. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).

62. Bay, S. K. *et al*. Trace gas oxidizers are widespread and active members of soil microbial communities. *Nat. Microbiol.* **6**, 246–256 (2021).

63. Wu, X., Cui, Z., Peng, J., Zhang, F. & Liesack, W. Genome-resolved metagenomics identifies the particular genetic traits of phosphate-solubilizing bacteria in agricultural soil. *ISME Commun.* **2**, 1–4 (2022).

64. Tully, B. J., Graham, E. D. & Heidelberg, J. F. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **5**, 170203 (2018).

65. Ortiz, M. *et al*. Multiple energy sources and metabolic strategies sustain microbial diversity in Antarctic desert soils. *Proc. Natl. Acad. Sci.* **118**, e2025322118 (2021).

66. Garner, R. E. *et al*. A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nat. Microbiol.* **8**, 1920–1934 (2023).

67. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

68. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).

69. Eddy, S. R. Accelerated Profile HMM Searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).

70. Ludwig, W. *et al*. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).

71. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).

72. van de Vossenberg, J. *et al.* The metagenome of the marine anammox bacterium 'Candidatus Scalindua profunda' illustrates the versatility of this globally important nitrogen cycle bacterium. *Environ. Microbiol.* **15**, 1275–1289 (2013).

73. Strous, M. *et al*. Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).

74. Stein, L. Y. Insights into the physiology of ammonia-oxidizing microorganisms. *Curr. Opin. Chem. Biol.* **49**, 9–15 (2019).

75. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing (2023).

76. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinforma. Oxf. Engl.* **30**, 2068–2069 (2014).

77. Bissett, A. *et al.* Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *Gigascience* **5**, s13742-016 (2016).

78. NEON (National Ecological Observatory Network). Soil microbe metagenome sequences. Dataset accessed from https://data.neonscience.org/data-products/DP1.10107.001.

79. Bahram, M. *et al.* Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).

80. Wilhelm, R. C. *et al.* A metagenomic survey of forest soil microbial communities more than a decade after timber harvesting. *Sci. Data* **4**, 170092 (2017).

81. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).

82. Brown, M. V. *et al.* Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative. *Sci. Data* **5**, 180130 (2018).

83. Hugerth, L. W. *et al.* Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biol.* **16**, 1–18 (2015).

84. Satinsky, B. M. *et al.* The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome* **2**, 17 (2014).

85. Biller, S. J. *et al.* Marine microbial metagenomes sampled across space and time. *Sci. Data* **5**, 180176 (2018).

86. Hu, W. *et al.* Metagenomics unravels differential microbiome composition and metabolic potential in rapid sand filters purifying surface water versus groundwater. *Environ. Sci. Technol.* **54**, 5197–5206 (2020).

87. Hendriksen, R. S. *et al.* Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* **10**, 1–12 (2019).

88. Bandla, A., Pavagadhi, S., Sridhar Sudarshan, A., Poh, M. C. H. & Swarup, S. 910 metagenome-assembled genomes from the phytobiomes of three urban-farmed leafy Asian greens. *Sci. Data* **7**, 278 (2020).

89. Mendes, L. W., Raaijmakers, J. M., de Hollander, M., Mendes, R. & Tsai, S. M. Influence of resistance breeding in common bean on rhizosphere microbiome composition and function. *ISME J.* **12**, 212–224 (2018).

90. Xu, J. *et al.* The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* **9**, 4894 (2018).

91. Howe, A. *et al.* Seasonal activities of the phyllosphere microbiome of perennial crops. *Nat. Commun.* **14**, 1039 (2023).

92. Levy, A. *et al.* Genomic features of bacterial adaptation to plants. *Nat. Genet.* **50**, 138–150 (2018).

93. Olson, D. M. & Dinerstein, E. The global 200: priority ecoregions for global conservation. *Ann. Mo. Bot. Gard.* **89**, 199–224 (2002).

94. Bivand, R. S., Pebesma, E. & Gomez-Rubio, V. *Applied Spatial Data Analysis with R, Second Edition*. (Springer, NY, 2013).

95. Bivand, R. *et al.* rgdal: bindings for the 'geospatial' data abstraction library. (2022).

96. Bivand, R. & Rundel, C. *Rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. (2020).

97. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* **46**, e59 (2018).

98. Matsen, F. A., Kodner, R. B. & Armbrust, E. V. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* **11**, 538 (2010).

99. Czech, L., Barbera, P. & Stamatakis, A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**, 3263–3265 (2020).

100. Woodcroft, B. J., Boyd, J. A. & Tyson, G. W. OrfM: a fast open reading frame predictor for metagenomic data. *Bioinformatics* **32**, 2702–2703 (2016).

101. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. VSURF: Variable Selection Using Random Forests. (2022).

102. Breiman, L., Cutler, A., Liaw, A. & Wiener, M. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. (2022).

103. Molnar, C., Bischl, B. & Casalicchio, G. iml: An R package for Interpretable Machine Learning. *JOSS* **3**, 786 (2018).

104. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2012).

105. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).

Acknowledgements

This work was supported by the Swedish University of Agricultural Sciences Senior Career Grant 2019-2025 to SH and the Swedish Research Council Grant Agreement No. 2016-03551 to SH. Computational resources were provided in part by Swedish National Infrastructure for Computing project SNIC 2022/22-1119 and NAISS 2023/22-1278, funded by the Swedish Research Council through Grant Agreement No. 2022-06725.

Author contributions

Conceptualization: S.H. Data curation: G.P., A.S. and C.J. Formal analysis: G.P. Funding acquisition: S.H. Investigation: G.P. Methodology: G.P., A.S., C.J. and S.H. Project administration: G.P. and S.H. Resources: S.H. Software: C.J., G.P and A.S. Visualization: GP, AS. Writing – original draft preparation: G.P., A.S., C.J. and S.H. Writing – review and editing: G.P., A.S., C.J. and SH.

Competing interests

The authors declare no competing interests.

Figure captions

Fig. 1: Comparative genomics indicates denitrifier types are unequally common and dominated by distinct phyla. **a** Overview of experimental approach for comparative genomics. Metagenome-assembled, single cell and isolate genomes derived from a variety of databases covering a broad range of environmental, host-associated, and engineered biomes [58–60,104]. After dereplicating genomes, denitrification genes were searched using a combination of HMM-based searches, phylogeny construction, and manual inspection of alignments. Maximum growth rate was inferred based on codon use bias in highly expressed ribosomal proteins. **b** Venn diagram of prevalence of denitrifier types. Reactions completed by each enzyme are placed on the outside of each circle, while the numbers inside each circle denotes the percentage of genomes among the 11,997 denitrifier assemblies with that genotype. Letters in italics denote whether genotype is considered an "initiator" (*i*), "terminator" (*t*) or both (*it*). **c** Stacked barchart denoting relative proportion of genomes belonging to phyla within each denitrifier type, with non-denitrifiers defined as containing neither *nir* nor *nosZ*, or being archaeal nitrifiers or anammox bacteria. **d** Proportion of genomes within dominant phyla encoding each denitrification enzyme gene, among denitrifiers. Phyla represented by fewer than 100 denitrifier genomes are aggregated under "other" and can be found in Supplementary Data 3.

Fig. 2: Estimated maximum growth rate of organisms based on codon usage bias. Stacked bar charts showing maximum predicted growth rate for genomes separated by denitrifier type (**a**) and NosZ clade (**b**). Slow-growing taxa (maximum growth rate < 0.2 h^{-1})

are depicted to the left on the horizontal axis and fast-growing taxa to the right. Genomes with fewer than 10 ribosomal proteins annotated were excluded from the analysis (n = 415).

Fig. 3: Relationship between denitrifier type relative abundance and resource availability in metagenomes derived from marine waters. a nitrate, **b** chlorophyll, and **c** chlorophyll:nitrate ratio. Relative abundance was normalized to the maximum prevalence observed for a given OTU prior to analysis to facilitate comparison between OTUs, and the slope between prevalence and nitrate, chlorophyll or chlorophyll:nitrate ratio was determined for each OTU. Nitrifiers are shown as triangles and all other as circles. Black squares with vertical lines denote weighted mean response and its 95% confidence intervals, respectively, where weighting is proportionate to inverse of the slope standard deviation. The "non-denitrifier" category includes genomes encoding none of the genes and/or just *nor*. The horizontal lines at zero denote the weighted mean response calculated using the inverse of slope standard error for each value. The two assemblies encoding the initiator-terminator combination Nir-Nos in this dataset are not shown.

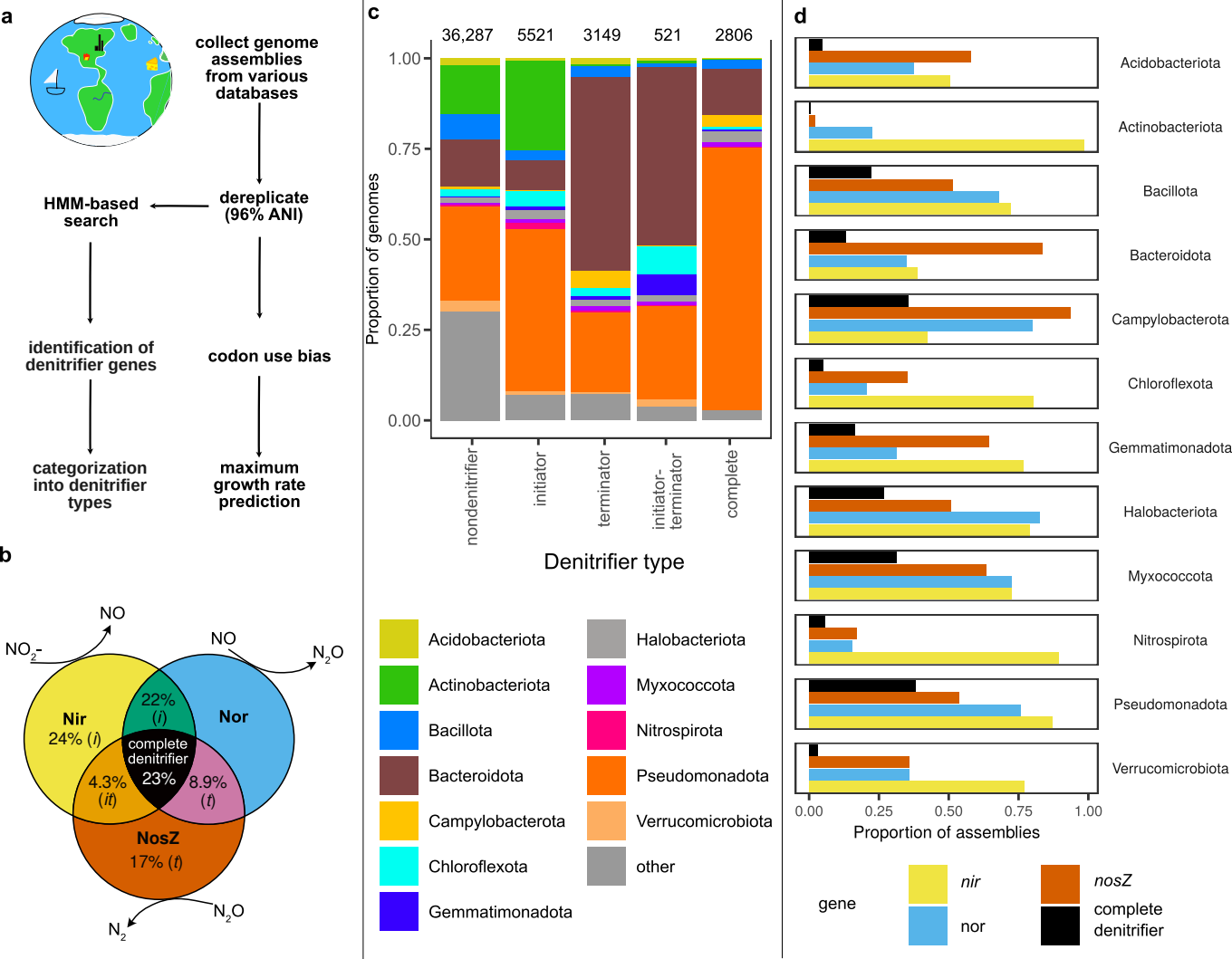
Fig. 4: Balance in prevalence of denitrification genes shows dominance of initiators at the community level across global biomes. Boxplots showing difference in *nos* versus *nir* counts (*nosZ* I + *nosZ* II) – (*nirK* + *nirS*); **a**) and difference in clade I and clade II *nosZ* counts counts (**b**) per Gb sequenced. Biomes were compared using Benjamini-Hochberg FDR-corrected pairwise ranked comparisons following Kruskall-Wallis, and common letters to the right in each boxplot denotes biomes with similar median gene prevalences. Numbers at right denote number of metagenomes included for each biome. Box boundaries represent first and third quartiles, with midline denoting median. Whiskers denote the less extreme of

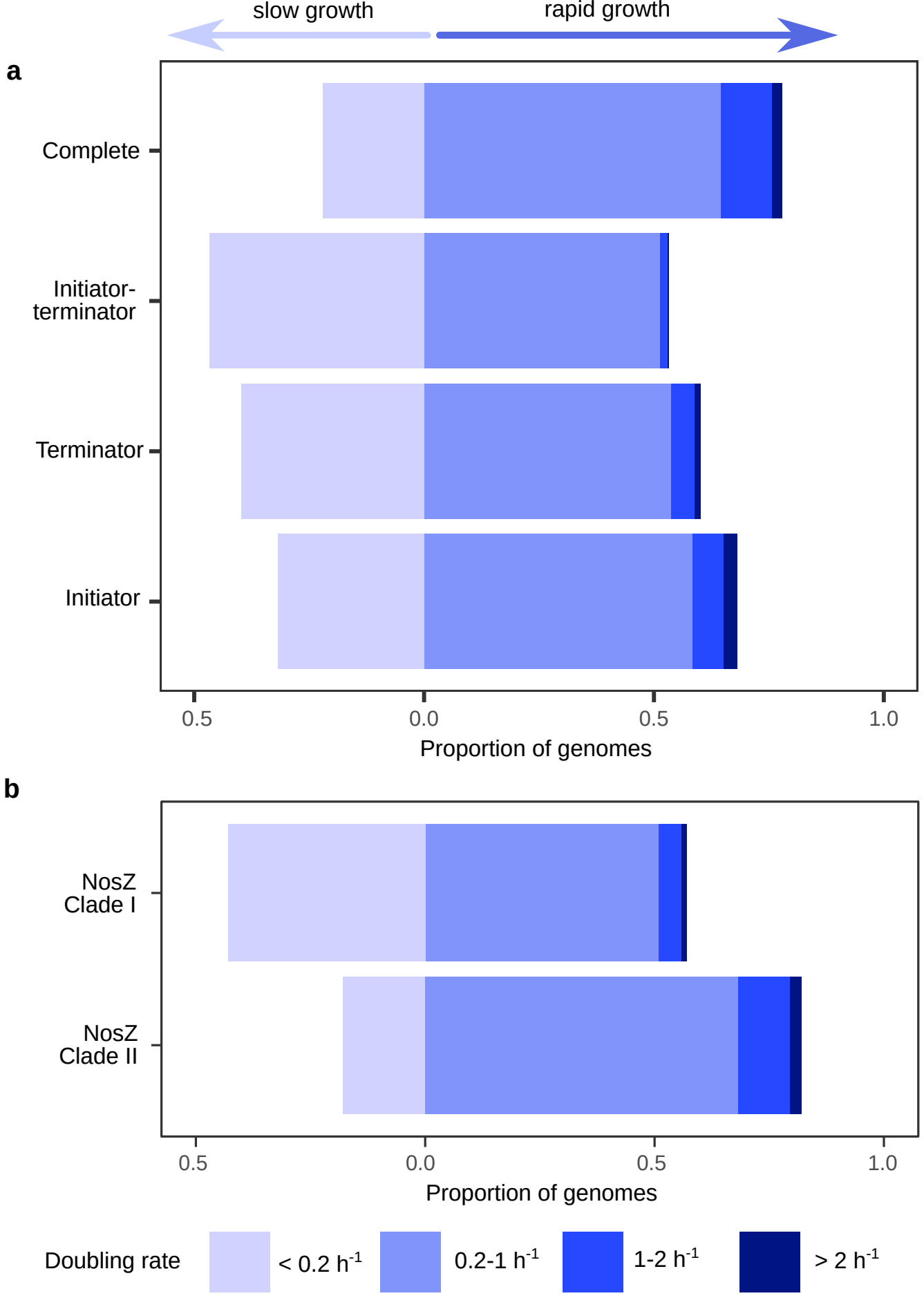
1.5 x interquartile range or minimum/maximum datapoint. **c** Mean phylum level composition of clade I and **d** clade II *nosZ* reads, organized by biome. Biomes represented by fewer than 20 metagenomes are excluded from the figure. The percentage of genomes having clade II *nosZ* and a complete denitrification pathway are indicated for each phylum in the legend and can be seen for both clades in Supplementary Fig. 2. Standard errors corresponding for each phylum are available in Supplementary Data 6. Host assoc. = host-associated; Engineer. = engineered.

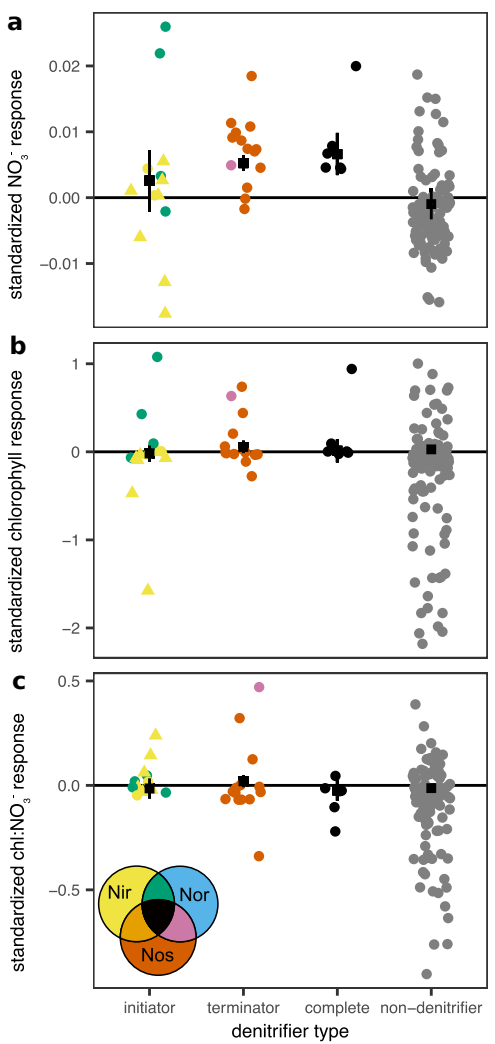
Fig. 5: Environmental predictors of balance between *nos* and *nir* counts using random forest models. The difference in *nos* and *nir* gene counts, calculated as (*nosZ* I + *nosZ* II) – (*nirK* + *nirS*), is normalized per Gb sequenced in soil (**a**) and marine (**b**) metagenomes. Abiotic predictors were used to generate accumulated local effects plots, which show the differences in prediction of the δ*nos-nir* (y-axis) compared to the mean prediction along the range of each predictor (x-axis), while accounting for potential correlations among predictor values. Values above zero indicate the model predicts higher than average dominance of *nosZ* over *nir* at a given value of the predictor variable. The analyses were performed on a subset of the soil (n = 298) and marine (n = 93) metagenomes for which relevant environmental metadata was available (see methods). Root mean square error and variance explained were 37 and 29% in the soil (**a**) and 20% and 20% in marine (**b**) models, respectively.

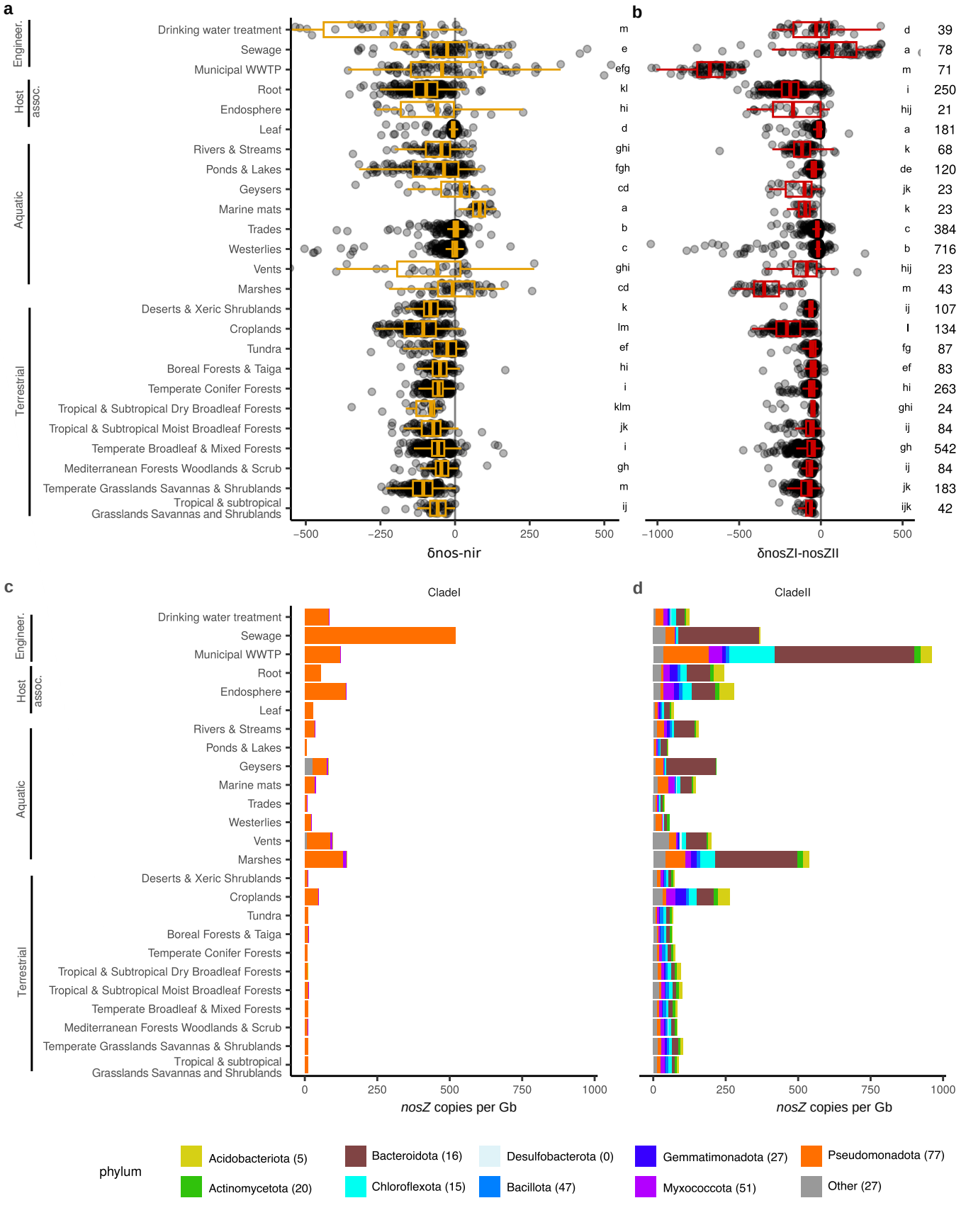
Fig. 6: Environmental predictors of balance between *nosZ* clade I and clade II using random forest models. The difference in *nosZ* clade I and II gene counts is normalized per Gb sequenced in soil (**a**) and marine (**b**) metagenomes. Abiotic predictors were used to generate accumulated local effects plots, which show the differences in prediction of the

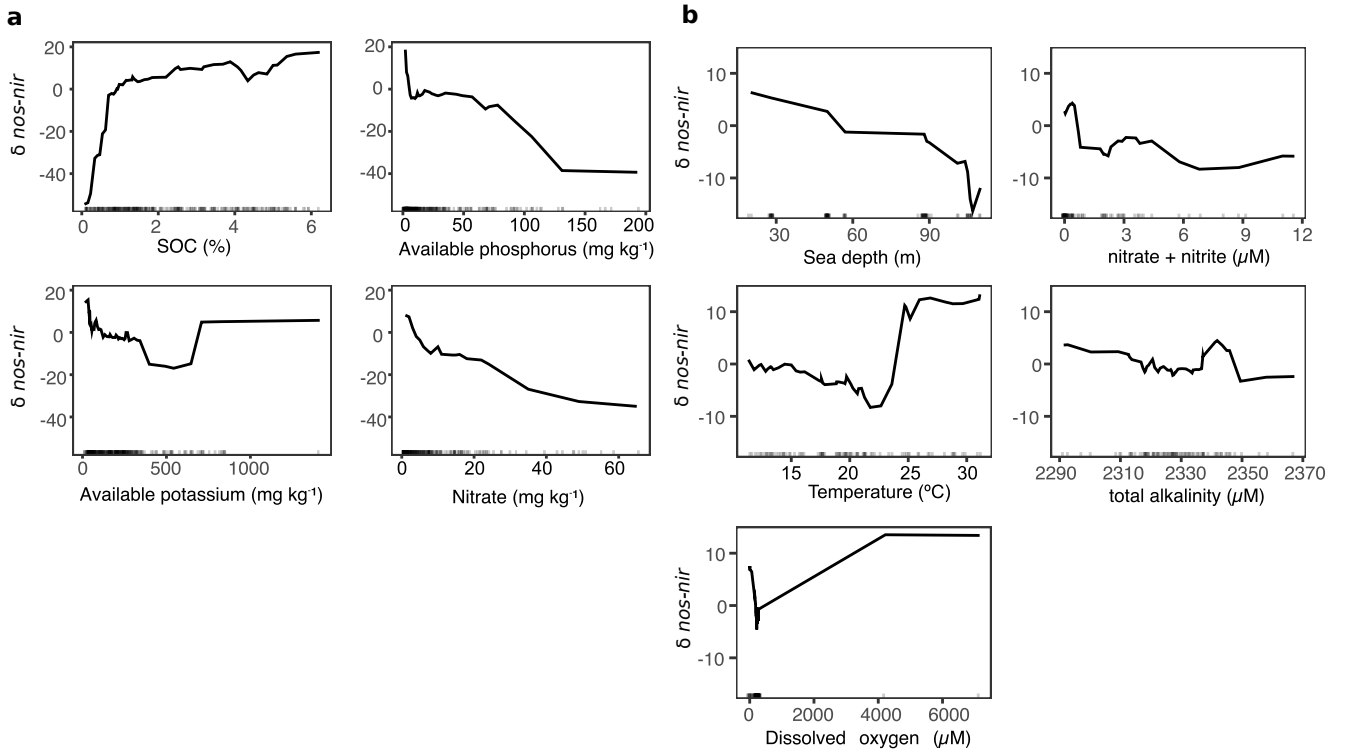
δ*nosZ*II- *nosZ*II (y-axis) compared to the mean prediction along the range of each predictor (x-axis), while accounting for potential correlations among predictor values. Values above zero indicate the model predicts higher than average dominance of *nosZ* clade I over *nosZ* clade II at a given value of the predictor variable. The analyses were performed on a subset of the soil (n = 298) and marine (n = 93) metagenomes for which relevant environmental metadata was available (see methods). Root mean squared error and variance explained were 40 and 59% in the soil (**a**) and 14 and 33% in marine (**b**) models, respectively.

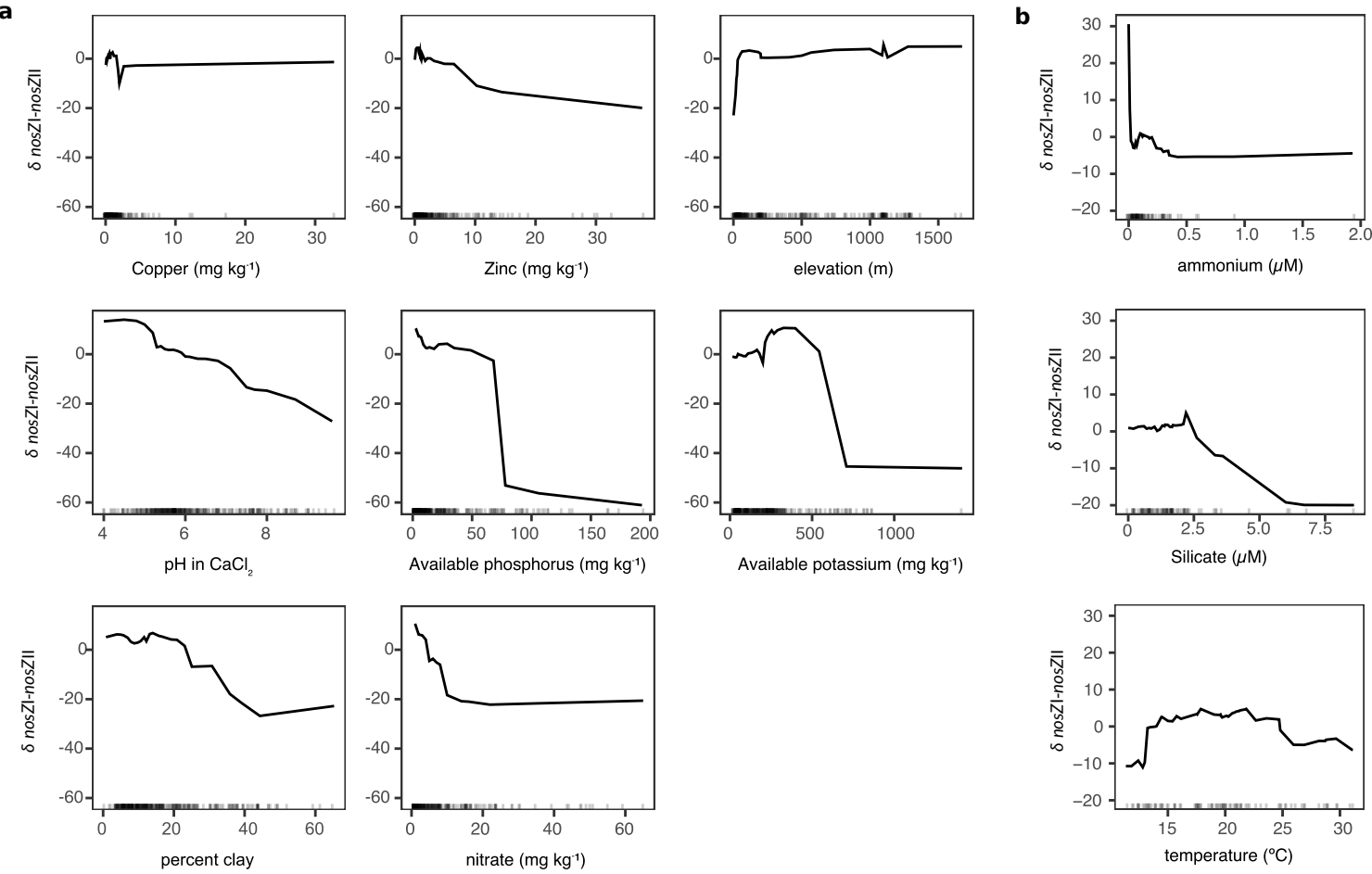

a
collect genome assemblies from various databases
dereplicate (96% ANI)
HMM-based search
identification of denitrifier genes
codon use bias
categorization into denitrifier types
maximum growth rate prediction
b
NO
NO2-
NO
N2O
Nir
24% (i)
22% (i)
Nor
complete denitrifier
23%
4.3% (it)
8.9% (t)
NosZ
17% (t)
N2O
N2
c
36,287
5521
3149
521
2806
Proportion of genomes
1.00
0.75
0.50
0.25
0.00
nondenitrifier
initiator
terminator
initiator-terminator
complete
Denitrifier type
Acidobacteriota
Actinobacteriota
Bacillota
Bacteroidota
Campylobacterota
Chloroflexota
Gemmatimonadota
Halobacteriota
Myxococcota
Nitrospirota
Pseudomonadota
Verrucomicrobiota
other
d
Acidobacteriota
Actinobacteriota
Bacillota
Bacteroidota
Campylobacterota
Chloroflexota
Gemmatimonadota
Halobacteriota
Myxococcota
Nitrospirota
Pseudomonadota
Verrucomicrobiota
0.00
0.25
0.50
0.75
1.00
Proportion of assemblies
gene
nir
nor
nosZ
complete denitrifier


slow growth
rapid growth
a
Complete
Initiator-
terminator
Terminator
Initiator
0.5
0.0
0.5
1.0
Proportion of genomes
b
NosZ
Clade I
NosZ
Clade II
0.5
0.0
0.5
1.0
Proportion of genomes
Doubling rate
< 0.2 h-1
0.2-1 h-1
1-2 h-1
> 2 h-1


a
standardized NO_3^- response
0.02
0.01
0.00
−0.01
b
standardized chlorophyll response
1
0
−1
−2
c
standardized chl:NO_3^- response
0.5
0.0
−0.5
Nir
Nor
Nos
initiator
terminator
complete
non−denitrifier
denitrifier type


a
b
c
d
Engineer.
Host assoc.
Aquatic
Terrestrial
Drinking water treatment
Sewage
Municipal WWTP
Root
Endosphere
Leaf
Rivers & Streams
Ponds & Lakes
Geysers
Marine mats
Trades
Westerlies
Vents
Marshes
Deserts & Xeric Shrublands
Croplands
Tundra
Boreal Forests & Taiga
Temperate Conifer Forests
Tropical & Subtropical Dry Broadleaf Forests
Tropical & Subtropical Moist Broadleaf Forests
Temperate Broadleaf & Mixed Forests
Mediterranean Forests Woodlands & Scrub
Temperate Grasslands Savannas & Shrublands
Tropical & subtropical Grasslands Savannas and Shrublands
δnos-nir
δnosZI-nosZII
CladeI
CladeII
nosZ copies per Gb
phylum
Acidobacteriota (5)
Actinomycetota (20)
Bacteroidota (16)
Chloroflexota (15)
Desulfobacterota (0)
Bacillota (47)
Gemmatimonadota (27)
Myxococcota (51)
Pseudomonadota (77)
Other (27)


a
b
δ nos-nir
SOC (%)
Available phosphorus (mg kg-1)
Available potassium (mg kg-1)
Nitrate (mg kg-1)
Sea depth (m)
nitrate + nitrite (μM)
Temperature (°C)
total alkalinity (μM)
Dissolved oxygen (μM)


a
b
δ nosZI-nosZII
Copper (mg kg-1)
Zinc (mg kg-1)
elevation (m)
ammonium (μM)
pH in CaCl2
Available phosphorus (mg kg-1)
Available potassium (mg kg-1)
Silicate (μM)
percent clay
nitrate (mg kg-1)
temperature (°C)