Denitrification is a community trait with partial pathways dominating across microbial genomes and biomes

Grace Pold [1], Aurélien Saghaï [2], Christopher M Jones [2], and Sara Hallin [2*]

[1] Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala

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

* Corresponding author: sara.hallin@slu.se

Running title: Global ecology of denitrification trait

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

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 pathway, but whether they dominate in natural communities and what metabolic traits and environmental factors drive the global distribution of complete vs. partial denitrifiers remains to be determined. To address this, we conducted a comparative analysis of denitrification genes in 61,293 genomes, 3,991 metagenomes covering all major biomes, and 413 terrestrial and aquatic metatranscriptomes. We show that partial denitrifiers outnumber complete denitrifiers and the potential to initiate denitrification is more common than the potential to terminate it, both among genomes and at the community level across most biomes, particularly in nutrient rich environments. These patterns were also reflected in the metatranscriptomes. Our results further indicate that complete denitrifiers are more likely to be fast-growing organisms, favoring organic acid over sugar metabolism, and encoding the ability to oxidize and reduce a broader range of organic and inorganic compounds compared to partial denitrifiers. This suggests complete denitrifiers are metabolically flexible opportunists. Together, our results indicate an environmental footprint on the presence of denitrification genes which favors the genomic potential for partial over complete denitrification in most biomes and highlight that completion of the denitrification pathway is a community effort.

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 (C) 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 [20,21], but also because the regulation of specific reactions may rely on intermediates that those executing the reaction themselves cannot use [22]. Further, there is less buildup of toxic intermediate compounds, which may allow microorganisms only capable of carrying out the initial steps of denitrification avoid the cytotoxic effects of NO_2^-, particularly under low pH environments [11,23]. This jointly indicates that split pathways may open more niches, thereby supporting a greater diversity of denitrifiers [21] . If organisms encoding partial vs. complete denitrification fulfill different niches, this should also be reflected in additional traits, such as growth rate, abiotic optima, and the ability to metabolize different substrates and use electron acceptors beyond those in the denitrification pathway. For instance, if complete denitrification is indeed associated with C influx, complete denitrifiers would be expected to be capable of higher growth and broader substrate catabolism compared to partial denitrifiers. Addressing the linkages between the complete or partial denitrification and the capacity for other traits can help identify the broader roles and conditions that favor each denitrifier type.

Here we combined comparative genomics (**Figure 1a**) with a global survey of denitrification gene ratios in environmental metagenomes and metatranscriptomes to evaluate the diversity of bacterial denitrifier types, their metabolic traits, 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, as well as via differential associations between denitrification enzyme genes and redox traits. Furthermore, considering organisms encoding a complete denitrification pathway are able to use a broader range of nitrogenous

terminal electron acceptors than those encoding fewer, we hypothesized that the former would have broader metabolic flexibility. We used a framework in which partial denitrifiers include genetically defined "initiators," which encode genes involved in NO_2^- but not N_2O reduction, and "terminators," which encode genes for N_2O but not NO_2^- reduction (**Figure 1b**). We used NO_2^- reduction, the decisive step of denitrification[24], as the first point of our analysis. We combined comparative genomics of 61,293 non-redundant bacterial isolate, metagenome assembled (MAGs), and single cell genomes with 3,991 metagenomes derived from all major biomes and 413 terrestrial and aquatic metatranscriptomes 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 complementary genomic traits associated with the denitrifier types in our framework.

Results

Partial denitrification dominates microbial genomes

We assessed the prevalence of partial and complete denitrification pathways among 61,293 bacterial genomes represented in GTDB release 214.1 (**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 [25]. Genomes carrying *nor* but not *nir* or *nosZ* accounted for 5% of the total number of assemblies and were excluded in downstream analyses because Nor in *nor*-only genomes are likely involved in detoxification rather than respiration. With this conservative approach, 18% of the

genomes (n=11,126) were considered denitrifiers and included *nir* and/or *nosZ* with or without *nor*. Among these, 23% encoded genes for complete NO_2^- reduction to N_2 ("complete denitrifiers"; **Fig. 1b**), with the complete denitrification trait more prevalent among isolates compared to MAGs (29% vs. 17%, respectively, **Supplementary Table 1**). However, there was considerable variation between phyla (**Fig. 1c, Supplementary Fig. 1**). Twenty-eight different bacterial phyla encoded the complete denitrification trait, with an overwhelming dominance of Pseudomonadota in our dataset (**Fig. 1c**). Among phyla represented by at least 100 denitrifier genomes, complete denitrifiers were most prevalent within Campylobacterota (38% of denitrifiers), Pseudomonadota (37%), and Myxococcota (26%, **Fig. 1d**).

Using our framework of initiators and terminators, we found that initiators encoding *nir* with or without *nor* but not *nosZ* were broadly more prevalent than terminators having *nosZ* with or without *nor* but not *nir* (49% vs. 24% of genomes with denitrification genes, respectively). This may represent the potentially divergent functions for NO_2^- versus N_2O reduction as a means of growth or detoxification vs. maintenance [26]. Partial denitrifiers occurred in 68 bacterial phyla, with initiators being overrepresented within Actinomycetota (98% initiators, 1% terminators) and Nitrospirota (88%, 8%) (**Fig. 1d**). However, Nitrospirota accounted for just a small fraction of initiator genomes (2.7%), while the majority were Pseudomonadota (47%) and Actinomycetota (22%) (**Fig 1c**). We found a higher proportion of terminators than initiators within Planctomycetota (76% terminators, 19% initiators) and Bacteroidota (60%, 18%) (**Fig. 1d**), with the latter phylum accounting for the majority of terminators in the dataset (53%; **Fig. 1c**). 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 13% of Gemmatimonadota, 13% of Chloroflexota, 9% of Bacteroidota and 5% of Verrucomicrobiota (**Fig. 1c**). 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 [27–30].

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 (74% in clade I vs. 26% in clade II; **Supplementary Fig. 2**), as previously observed [4]. Within clade II, complete denitrification was most prevalent in Pseudomonadota and Aquificota, with 80% of Pseudomonadota and 94% of Aquificota encoding complete denitrification compared to an average of 20% in the remaining phyla. This suggests that the difference in clade I vs. clade II prevalence can be used as a proxy for complete vs. partial denitrifier prevalence within the N_2O reducing community, but not if clade II is dominated by Pseudomonadota or Aquificota.

Inferred ecological preferences of complete vs. partial denitrifiers

We next assessed whether the genetic potential for complete denitrification or for initiating rather than terminating denitrification was associated with the predicted growth rate and resource use patterns among bacteria. Using the codon bias-based maximum growth rate prediction tool gRodon [31], we categorized genomes into fast- (growth rate faster than 0.2 h^{-1}) and slow-growing (growth rate lower than 0.2 h^{-1}) [31]. Rapid growth was more common among complete denitrifiers (83%) compared to initiators (70%), terminators (65%) and initiator-terminators (62%; **Fig. 2a upper**

panel) and for genomes encoding clade I (85%) compared to clade II NosZ (65%; **Fig. 2a lower panel**). Similarly, estimated median growth rates of complete denitrifiers (0.46 h^{-1}) were greater than that of initiators (0.28 h^{-1}), terminators (0.23 h^{-1}) and initiator-terminators (0.20 h^{-1}; overall Kruskall-Wallis $\chi^2 = 41$, Dunn $P < 0.001$ in all cases). This faster potential growth rate was associated with a lower sugar to acid preference of complete denitrifiers compared to all three of the categories of partial denitrifiers, based on the total sum of genes involved in each pathway [32] (**Fig. 2b**). This was accompanied by a higher density of transcription factors (**Fig. 2c**) and transporters (68 Mbp^{-1} in complete denitrifiers vs. 65 Mbp^{-1} in initiators, 40 Mbp^{-1} in terminators, and 42 Mbp^{-1} in initiator-terminators, $P < 2 \times 10^{-16}$). Furthermore, complete denitrifiers were inferred to grow on more of the 56 organic substrates examined (median of 20 corresponding to 4.75 Mbp^{-1}), compared to initiators (13; 3.32 Mbp^{-1}), terminators (11; 2.90 Mbp^{-1}), and initiator-terminators (9; 2.55 Mbp^{-1}; **Fig. 2d**). This was driven by complete denitrifiers using a broader range of organic and amino acids compared to partial denitrifiers (median of 15 vs. 6-7), rather than sugars (median of 2 vs. 2-3).

We subsequently examined the potential for partial and complete denitrifiers to use a range of inorganic compounds as electron donors and acceptors (**Fig. 2e**). While the ability to oxidize inorganic compounds such as hydrogen may be associated with an ability to tolerate energy-limited conditions [33], the ability to reduce inorganic compounds may indicate that one type or another of denitrifiers is better able to tolerate lower redox conditions. Although most genomes lacked the genetic potential to oxidize inorganic substances or reduce non-nitrogenous terminal electron acceptors (median = 0), complete denitrifiers had the genetic potential to both oxidize and reduce

more compounds than partial denitrifiers (Kolmogorov-Smirnov statistic $P < 0.0001$ in all cases).

The inferred abiotic niche breadths of complete denitrifiers were not uniformly broader than that of partial denitrifiers. Complete denitrifiers had slightly broader inferred temperature ranges compared to initiators (26.84 vs. 25.12 ^{o}C; t = -24.583, $P < 2 \times 10^{-16}$), terminators (25.54 ^{o}C, t = -18.569, $P < 2 \times 10^{-16}$) and initiator-terminators (25.93 ^{o}C, t = -8.038, $P = 1.02 \times 10^{-15}$). pH ranges were narrower in complete denitrifiers compared to initiators (3.75 vs. 3.83; t = 8.136, $P = 4.44 \times 10^{-16}$) and broader compared to terminators (3.58; t = -6.756, $P = 1.50 \times 10^{-11}$). Salinity ranges were broader in complete denitrifiers than initiators (5.64 vs. 5.31; t = -3.775, $P = 0.000162$) but narrower than in initiator-terminators (5.92; t = -2.441, $P = 0.014665$).

We subsequently assessed the degree to which complete and partial denitrifiers were associated with high or low resource availability using CoverM [34] to map reads from the TARA Oceans [35] project to MAGs derived from the same data [36]. No subset of the metagenome data we collected or other databases we screened had a sufficient number of complete denitrifier MAGs and environmental data to allow for this analysis, highlighting the rarity of complete denitrifiers in global ecosystems. Nitrate was used as a proxy for denitrification electron acceptor availability and chlorophyll concentration as a proxy for electron donor supply because of its positive association with gross primary productivity [37]. The proportion of reads mapping to the collection of 170 MAGs in the dataset differed by sample and ranged from 5 to 30%. We calculated a standardized environmental response 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 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 (**Supplementary Fig. 3a**). Mean standardized chlorophyll and chlorophyll: NO_3^- ratio responses were similar across all groups of organisms, independent of denitrifier type (**Supplementary Figs. 3b,c**). One possible reason for the apparent lack of correlation between denitrifier type and resource availability is that these bulk measurements likely do not represent the microhabitat that denitrifiers experience. For instance, pre-filtering of the samples depleted much of the particulate organic matter, where denitrification activity is concentrated [38], and where modeling indicates that complete denitrifiers are favored [21]. Nonetheless, the results from this pelagic marine dataset do not indicate that measured resource availability affects the prevalence of complete or partial denitrifier genomes.

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 using an HMM-based search and phylogenetic placement approach [39]. The difference in the copy number of *nosZ* and *nir* genes normalized to Gbp sequenced in each metagenome (δ*nos-nir*) was used as a proxy for the genetic potential for termination compared to initiation of denitrification, and *nosZ* clade I vs. clade II (δ*nosZ*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 cases 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*, indicating a greater or similar potential to initiate than terminate denitrification within the denitrifier communities (**Fig. 3a**). Marine mats served as an exception, and *nosZ* prevalence exceeded that of *nir* by 60%. This would occur if there were higher rates of NO_3^- assimilation or low inputs of NO_2^- from nitrification [40], as suggested by the overall low prevalence of *nir* and lack of ammonifier associated *nirK* clades observed previously in these samples [25]. *nosZ* Clade II was more prevalent than *nosZ* Clade I in nearly all biomes considered, (**Fig. 3b**). This indicates that partial denitrifiers generally dominate terminator communities because partial denitrification is more prevalent among Clade II *nosZ*. This is further supported by our observation that phyla depleted in complete denitrifiers, such as Bacteroidetes, Chloroflexota and Gemmatimonadota, dominated *nosZ* clade II across biomes (**Figs. 3c,d**). Therefore, we can conclude that the majority of organisms capable of terminating denitrification in global biomes do not also initiate this process. This is particularly the case in croplands, marshes, and activated sludge from municipal wastewater treatment plants, which were the biomes where δ*nosZ*I-*nosZ*II is lowest for terrestrial, marine, and engineered environments (**Fig. 3b**). The dominance of *nosZ*II coincides with the highest total *nosZ* and *nir* gene abundances (**Supplementary Data 1**), indicating conditions that are overall more favorable for denitrification also 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 *nir* gene fragment prevalence, and seems to largely reflect the preponderance of *Acidovorax* in these samples [41].

We also assessed the correlation between the diversity of denitrifiers based on phylogenetic diversity of the dominant nitrite reductase gene, *nirK* and potential initiator-terminator or clade I and II differential abundance [25]. We found that *nirK* phylogenetic diversity was positively correlated with δ*nos-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 *nirK* diversity was positively correlated with δ*nosZ*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 capacities for initiating and terminating denitrification are relatively more balanced, such that complete denitrification at the community level is 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 [42]. We generated accumulated local effect plots to show the main effect of the target variable on predicted *δnos-nir* (**Fig. 4**) and δ*nosZ*I-*nosZ*II (**Fig. 5**), while accounting for the other predictors. Models explained less than half of the variance in gene differences, except for δ*nosZ*I-*nosZ*II in soils. Increases in soil organic carbon content up to ~1% positively affected δ*nos-nir*, while the effect of soil NO_3^- content was always negative (**Fig. 4a**). *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 the sum 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 ^{o}C (**Fig. 4b**).

Available NO_3^-, phosphorus, potassium, and zinc had a negative effect on predicted δ*nosZ*I-*nosZ*II (**Fig. 5a**), which suggest that partial N_2O reducers increase with increasing nutrient levels. pH and clay content also had a negative effect on predicted δ*nosZ*I-*nosZ*II, though only pH had a negative effect over the entirety of measured values. Increases in elevation at low altitude were associated with increased predicted δ*nosZ*I-*nosZ*II, but additional increases in elevation were not associated with further increases. In the marine study, temperature and ammonium concentrations had opposing unimodal effects on predicted values over their observed ranges (**Fig. 5b**).

Expression of denitrification termination and initiation in the environment

We subsequently quantified the presence of *nosZ* and *nir* transcripts in 413 metatranscriptomes from soils and aquatic environments, distributed among five biomes, to assess whether the biome-level differences and associations between environmental variables and δ*nosZ*I-*nosZ*II and δ*nos-nir* gene abundance were also apparent in gene expression. *nirK* transcripts were strongly dominated by archaeal nitrifier reads (median 50%, range 0-100%), which we excluded in our analysis (see methods). Consistent with the metagenome analysis, *nir* was more prevalent than *nosZ* and clade II *nosZ* was more prevalent than clade I *nosZ* (**Fig. 6a**). The soil studies either directly or indirectly manipulated carbon availability, and soils associated with higher carbon availability had higher δ*nos-nir* expression than their paired lower-carbon samples (SMD 0.75, p=0.015) but did not have higher δ*nosZ*I-*nosZ*II (SMD 0.06, p = 0.73). Among aquatic studies, which were all observational, correlation

coefficients were negative between δ*nos-nir* and the NO_3^- + NO_2^- (mean -0.27, CI: -0.53 - 0.00) and positive between δ*nos-nir* and chlorophyll concentrations (0.22 CI 0.04, 0.41; **Fig. 6b**). δ*nos-nir* was not correlated with dissolved oxygen or phosphate concentrations, or with bacterial production, another proxy for resource availability. δ*nosZ*I-*nosZ*II was positively correlated with dissolved oxygen content, which would occur if the dominant organisms encoding clade I NosZ lowered oxygen concentrations around their N_2O reductase better than those encoding clade II [43].

Discussion

By combining a comparative analysis of genomes with broad metagenomic and metatranscriptomic surveys, 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 of MAGs from different environments [5,9,44–47], 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. Our results show denitrification is primarily a community trait based on division of labour, thereby relying cross feeding between microorganisms producing and consuming NO_2^-, NO and N_2O to complete the pathway. Furthermore, other genes and genomic traits implicate complete denitrifiers as metabolically flexible generalists compared to any partial denitrifier type.

Various explanations exist for the dominance of partial denitrifiers. We show that partial denitrification enables greater overall niche partitioning, which was supported by narrower substrate ranges, lower capacity for oxidizing and reducing various inorganic electron donors

and acceptors, and a lower genomic density of transcription factors and transporters among partial denitrifiers compared to complete denitrifiers. In the environment, higher *nirK* phylogenetic diversity, an indication of greater functional diversity among nitrite reducers [25], was associated with an increasing balance between capacity to initiate and terminate denitrification at the community level. This would further support niche partitioning if the pathways were split among partial denitrifiers, which is likely considering that that we mainly observed negative δ*nos-nir* values in the environment and the fact that MAGs are predominantly (84%) partial denitrifiers. Partial denitrifiers may also dominate communities due to the presence of a rate-efficiency tradeoff, i.e. that energy flux is slower through long pathways but potentially enables more complete usage of the terminal electron acceptor when it is limiting [19]. However, the dominance of partial denitrifiers among N_2O reducers in the environment was not coherently explained by a rate-efficiency tradeoff. If this were the case, we would expect complete denitrification to decline as NO_3^- increases or as C:NO_3^- and C:NO_2^- ratios decrease, assuming denitrifiers are predominantly heterotrophs in the environment. This agrees with predictions based on modelling and observations in oxygen minimum zones in the oceans, proposing that the prevalence of a complete pathway increases as the limiting substrate shifts from C to N [21]. Accordingly, our random forest modeling inferred that the N_2O reducing community in higher NO_3^- soils became more dominated by partial denitrifiers, i.e. those carrying *nosZ* clade II, and fertilized croplands had more negative *δnosZ*I-*nosZ*II than other soils in our cross-biome study. However, a similar reduction in *δnosZ*I-*nosZ*II was not associated with increasing NO_3^- + NO_2^- in aquatic metagenomes and metatranscriptomes or in the analysis mapping reads to marine MAGs. In addition, there was no relationship between increasing carbon content in soil or higher chlorophyll concentrations in aquatic samples and higher prevalence of complete denitrifiers in either metagenomes or metatranscriptomes. This is consistent with a recent qualitative

analysis of ~1600 MAGs [46], which concluded that a rate-efficiency tradeoff cannot explain observed partial denitrifier dominance. Based on our findings of broader substrate ranges and greater genomic allocation to substrate uptake and transcriptional regulation, we instead propose that complete denitrifiers are adapted to flexibly take advantage of resources varying in time and space, while partial denitrifiers are restricted to slower and therefore less variable growth rates. This may explain the over-representation of complete denitrifiers among model denitrifiers and isolates that grow readily under standard resource-rich lab conditions. Under carbon-rich conditions supporting rapid growth in the environment, complete denitrifiers may benefit from using available NO_3^-/NO_2^- more efficiently [21], while under rapidly fluctuating conditions many have the flexibility to use whichever electron donor and terminal electron acceptors are most readily available. This indicates that partial denitrifiers dominate because they have cheaper metabolisms to run, that do not require such complex regulatory mechanisms. This is consistent with the overall dominance of slow-growing organisms in natural environments [31,48,49].

The community level genetic balance between initiation and termination of denitrification indicates a higher capacity for N_2O production than reduction in nearly all biomes. The variations between biomes and among the specific ecosystems within biome categories were readily explained by environmental variables. We found that high NO_3^- levels were associated with lower *δnos-nir* in both soils and aquatic systems, indicating greater imbalance between the two genes and a decrease in the relative importance of N_2O reducers. In soils, the same pattern was observed for phosphorous. High concentrations of NO_3^- may also allow denitrifiers to outcompete ammonifiers, which both theoretical [50,51] and empirical evidence indicate prevail under high $C:NO_3^-$ [52,53], including in the soil metagenomes used in this study [54]. Overall, this shows that nutrient rich environments promote denitrifier communities

dominated by initiators and similar patterns were observed at the transcriptional level in aquatic environments (no data for soils). The observed imbalance in the denitrifier community could explain why nutrient loaded environments like agricultural soils are major sources of N_2O. However, complete denitrifiers have been observed to express only enzymes involved in denitrification initiation under high NO_3^- and/or NO_2^- [55,56], which also lead to accumulation of N_2O. Furthermore, high fertilization rates commonly favor nitrifier-denitrification [57], and hydroxylamine from nitrification and high NO_2^- concentrations can both induce N_2O production from chemodenitrification [57,58]. Finally, N_2O may also be produced during the detoxification of NO from host immune responses [59,60]. Therefore, there are a wealth of pathways that are distinct from denitrification and can reduce NO_2^- and/or produce N_2O, and it is plausible that inter-pathway competition for NO_2^- favors the presence of initiators ready to consume NO_2^- in denitrification rather than letting it enter other pathways. Terminators could conceivably take advantage of the accumulated N_2O from all these possible sources, in particular obligate terminators (i.e. predominantly *nosZ* clade II) which use N_2O as terminal electron acceptor and thereby serve as important N_2O sinks [17,61]. This reasoning agrees with the observed low δ*nosZ*I-*nosZ*II in metagenomes from agricultural soils and wastewater treatment plants. We conclude that high levels of NO_3^- allow both initiation and termination specialists to prevail, but with a dominance of initiators which altogether increase the risk of N_2O emissions.

Fixed carbon availability was associated with reduced dominance of *nir* over *nosZ* in aquatic transcriptomes and soil metagenomes, in particular in soils with less than 1% SOC. The soil metatranscriptomes also indicated increased carbon availability could help Nos*Z* overcome a low competitiveness for electrons [62–64]. However, the chemical composition of the available carbon is expected to affect the balance between initiation and termination of denitrification

and complete vs. partial denitrifiers. Organic acids have short catabolic pathways and fewer opportunities for enzyme bottlenecks to slow the flow of electrons into the electron transport chain compared to sugars and should favor complete denitrifiers able to use an array of terminal electron acceptors, in agreement with our comparative genome analysis. At the community level however, organisms metabolize and co-metabolize [65] substrates from among tens of thousands of molecularly-distinct potential electron donors of varying bioavailability [66,67], leading to conflicting effects of sugar and organic acids addition on N_2O to N_2 production from soils [68–70]. High dissolved oxygen concentrations were also positively associated with dominance of clade I over clade II *nosZ* I expression but not gene abundance in aquatic samples. There is no clear consensus on clade I being less sensitive to high oxygen than clade II NosZ [71,72], and oxygen niche of N_2O reducers may be a function of organism oxygen consumption [43] or even indirectly their competition for NO_2^- with nitrite oxidizing bacteria [21]. Thus, while the mechanism for oxygen preference cannot be inferred from our data, our results nonetheless indicate that terminators and complete denitrifiers make distinct contributions to N_2O reduction as a function of environmental conditions.

Our results indicate that the prevalence of denitrification functional types 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 environmental conditions, but also through biotic interactions. Denitrification does not occur in isolation, and its steps may occur in response to its intermediates being produced by other pathways. Further, spatiotemporal variation in electron donor and acceptor ratios can enable coexistence of organisms completing different steps of denitrification. Future work addressing phenotypic plasticity of denitrifiers under realistic environmental conditions is necessary to resolve this uncertainty and disentangle biotic and abiotic drivers of denitrifier community assembly. This

will be particularly enhanced by improved generation of high-quality assemblies of representative environmental genomes, and by the ability to sample microbial communities on the scale at which they interact with one another and the environment.

Methods

Denitrification enzyme databases

Alignments and phylogenies for NirK and NirS were obtained from Pold *et al.* [25]. The Nor reference database was derived from Murali *et al.* [73] and consisted of 67 Nor sequences (cNOR, qNOR, bNOR, eNOR, gNOR and sNOR, and nNOR) and 865 other heme copper oxidases. 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.* [25]. After manually checking the alignment for conserved residues [74] and trimming the alignment to its conserved core, we built a phylogeny using IQ-TREE v. 2.1.3 with the best model identified by modelfinder (LG+R10) [75]. 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 [24,47,76].

Genome annotation

We used the representative set of genomes from GTDB v214.1 for our analysis (**Supplementary Data 2**).Genomes less than 80% complete, more than 5% contaminated, or which failed taxonomic consistency of contigs based on GUNC v.1.0.6 [77] were excluded from our analysis, resulting in 61,293 genomes in our final dataset. We included both isolate (49% of assemblies) and culture-independent genome assemblies (metagenome assembled and single cell genomes; 51%) in our analysis to

balance genome quality with inclusion of genetic diversity representative of environmental samples. We identified genes for NirK, NirS, Nor, and NosZ in each genome using an HMMsearch [78] against the corresponding amino acid databases as described above. After aligning potential proteins to the reference, we imported them into ARB v.7.0 [79] and used a combination of phylogeny building with FastTreeMP v.2.1 [80] 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 *nosZ*), and partial denitrifiers (lacking *nir* or *nosZ*), with partial denitrifiers further divided into "initiators" (*nir* with or without *nor*), terminators (*nosZ* with or without *nor*) and initiator-terminators (*nir* and *nosZ*; **Fig. 1b**). Anammox genomes carrying *nir* (n = 47) were classified as non-denitrifiers because their nitrite reductase is not thought to be involved in denitrification [81]. Similarly, genomes encoding just *nor* were denoted as non-denitrifiers (n = 3,064) as a conservative measure since they are likely involved in detoxification rather than respiration. Since we only used bacterial genomes, our analysis includes ammonia oxidizing bacteria potentially capable of nitrifier-denitrification, but not ammonia oxidizing archaea for which a respiratory function for nitrite reductase has not yet been established [82,83].

A range of tools were used to identify and validate genes associated with potential electron donors and acceptors used by microorganisms. Genes for proteins involved in arsenite oxidation (AioBA; validated with alignment from Quemeneur *et al.* [84]), anammox (Hzs, Hdh) and iron redox were identified in genomes using MagicLamp v.1.0 with final annotations

curated to match minimum subunit composition proposed for each enzyme [85]. We used AmoA as a marker for ammonia oxidation in bacteria using the alignment of AmoA, HmoA and PmoA from Diamond *et al.* [5]. Respiratory reductive dehalogenases were identified using the RDaseDB alignment (v.2020 [86]) combined with a literature search to exclude catabolic enzymes; we note that the TAT motif proposed to be indicative of respiratory enzymes [87] was absent from the majority of sequences and could not be used. Sulfur redox genes were identified using HMSS2 and categorized into oxidation or reduction-associated enzymes following Tanabe and Dahl [88]. Sulfide: quinone reductase (SQR) enzymes were excluded from our analysis, both because their physiological role is variable, and because we found the HMMs were unspecific. Hydrogenases were identified using MagicLamp, refined using the residues in Greening *et al.* [33], and verified as to their clade and probable physiological function using the HydDB web server [89]. Clades 1a-d, 1g,h and j and clade 2 were considered involved in hydrogen oxidation, and clades 4 a,b,c and h were considered involved in reduction. ArrA and ArxA involved in arsenate reduction and arsenite oxidation, respectively, were identified using an HMM built from sequences from Wells *et al.* [90], and validated using the residues from Ospino *et al.* [91]. All redox gene annotations are reported in **Supplementary Data 3**.

Sugar-acid preference of organisms was inferred using the sum of KEGG orthologues identified by enrichM (v0.6.3, default settings) that were annotated to sugar vs. acid metabolism [32], noting that the majority of the genomes fall outside the originally-calibrated range of GC contents (**Supplementary Data 4**). GapMind carbon [92] (March 22nd, 2021 database version) was run to infer ability to grow on 62 compounds. We used custom pathway completeness cutoff scores for each pathway based on annotations for organisms for which growth had been reported on that substrate in BacDive (**Supplementary Data 5**). Our

positive controls dataset did not have data for 6 substrates, leaving 56 for analysis. Transporters were annotated using DIAMOND against the TCDB database, with hits showing at least 40% identity to the reference and 70% coverage of both reference and query sequence, with no more than 10% difference in alignment length kept for analysis [93]. Transcription factors were annotated using DeepTFactor [94] (v. 2020-11-09) with default settings (**Supplementary Data 6**). We used GenomeSPOT to infer temperature, pH, and salinity ranges [95], and used the difference between the maximum and minimum values for each genome in our analysis (**Supplementary Data 7**).

Unless otherwise noted, all subsequent analyses were completed in R v.4.1 [96]. The maximum growth rate of each taxon was estimated using gRodon [31], which used codon usage bias of ribosomal proteins annotated in the genomes using hmmearch against the ribosomal proteins extracted from Uniprot PFAM v. 35.0 with profile-specific cutoffs (**Supplementary Data 8**). Differences in growth rates between denitrifier types were assessed using Kruskall-Wallis tests (stats::kruskal.test) followed by Dunn tests with Benjamini-Hochberg correction for multiple testing (FSA::dunnTest), 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 [31].

To determine whether complete and partial denitrifiers differ in transcription factor, transporter, carbohydrate-active enzyme, or catabolized substrate count or sugar acid preference, we fit linear models, logging or taking the exponent of variables where needed. However, we analyzed GapMind data with a generalized linear mixed model with a Tweedie distribution in glmmTMB[97] (v.1.1.10) due to the poor fit of OLS and

high proportion (10-37%) of assemblies where no substrates within a chemical class were predicted to be used. We subsequently ascertained that the model residuals were not phylogenetically clustered using phytools::phylosig (v.2.3.0 [98]) and an organism tree derived from the GTDB reference tree [99]. This tree was rerooted in Fusobacteriota [100] and forced to be ultrametric using the castor::date_tree_red function [101] (v.1.8.2). Blomberg's K [102] of model residuals was less than one in all instances ($P < 0.05$), so we did not include phylogenetic relatedness as a random effect in our analysis. We evaluated differences between complete and partial denitrifier genome content using post-hoc multiple comparisons with the sandwich::vcovHC function (v.3.1.1 [103]) to account for heteroskedasticity. Due to the large number of zeroes, we assessed the hypothesis that partial denitrifiers encode the potential to oxidize and reduce a smaller number of inorganic compounds than complete denitrifiers by comparing the cumulative density distributions using a one-sided Kolmogorov-Smirnov test using stats::ks.test.

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 (**Supplementary Data 9**). The majority of soil metagenomes came from the Australian Microbiome Project/BASE [104], the National Ecological Observatory Network [105], Topsoils Microbiome Project [106], Long-Term Soil Productivity experiment [107], and the Stordalen Mire [108]. The majority of aquatic metagenomes came from the Australian Microbiome Project [109], Linnaeus Microbial Observatory [110], the Amazon Continuum Project [111] and BATS, GEOTRACE and HOTS [112]. Engineered metagenomes included those from

drinking water [113] sewage [41], and wastewater treatment plants [114]. The host-associated metagenomes dataset consists primarily of plant-associated sequences, such as from leafy greens [115], beans [116], citrus rhizosphere [117], switchgrass [118] and *Arabidopsis* [119], 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.

We completed a similar search for metatranscriptomes, focusing on soil (including rhizosphere) and marine studies where carbon and/or nutrient availability had either been manipulated or logic or metadata indicated they differed between samples (**Supplemental Data S10**). These metatranscriptomes were identified based on searching the American Society for Microbiology journals website for metatranscriptomes and google databases for the terms “soil” or “marine” and “carbon” or “nutrient” and “metatranscript*”.

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 150 nt 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. Our final set of metatranscriptomes only shares samples with a handful of samples from a single study in our metagenome dataset and contains 413 metatranscriptomes (64 rhizosphere or soil with root exudates added, 92 other soil, 185 primarily coastal and estuarine marine, 24 lake sediment, and 48 river metatranscriptomes).

Biomes were assigned to metagenomes based on The Nature Conservancy Terrestrial Ecoregions for terrestrial samples [120] 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 [121], rgdal v.1.523 [122], and rgeos v0.5-5 [123]. 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 [39] to identify denitrification genes (*nirK*, *nirS* and *nosZ*) in the metagenomes and metatranscriptomes, as previously described [25]. 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; [124]). We used the accumulate function in gappa[125] 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 (median 0% of *nirS* reads in metagenomes and 0% in metatranscriptomes), archaeal nitrifier NirK clades 2 and 4 (1.4% of *nirK* reads in metagenomes and 50% in metatranscriptomes), as well as the eukaryotic NirK clades 1b and 1e (0% and 0%) [25]. Our analysis also excluded the 0.4% of reads placed in the bioinformatically inferred “clade III” *nosZ* [10], which forms the outgroup in our reference tree. This metagenome and metatranscriptome search depended on: hmmer (v.3.2.1; [78]), OrfM (v.0.7.1; [126]), 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 included homologous multicopper oxidase proteins from Cyanobacteriota and Thermoprotetoa, NirN and NirF, and various yet to be characterized NosZ-like proteins, respectively. 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 metatranscriptomes we standardized based on Gbp mRNA sequenced, which was determined by excluding rRNA reads from total RNA using SortMeRNA v.4.3.7 with the smr_v4.3_default_db database and settings "--other --fastx --num-alignments 1 --no-best". For *nir* vs. *nosZ* abundance we used Eq. 1 such that positive values denote *nosZ* is more abundant than *nir*:

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

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

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

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 and 26% of metatranscriptomes where *nosZ*I was not detected but *nos*ZII was found in 1-13,041 (1-10,926 in metatranscriptomes) 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 [35] 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; [34]; -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 gene counts and resource availability using random forests. This analysis was restricted to metagenomes from the BASE and Australian marine projects[104,112] 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 2 and 3). We subsequently ran VSURF v. 1.1.0[127] 100 times to identify the best predictors for each ratio and biome, keeping only those variables retained in at least 95 of the 100 iterations. We then used the randomForest package v. 4.7.1-1[128] to model the relationship between gene ratios and the final set of retained variables. 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 [129].

Finally, we verified whether high carbon and/or low nutrient availability favored the transcription of *nosZ* over *nir* or clade I *nosZ* over clade II *nosZ* by calculating $\delta nos-nir$ and $\delta nosZ$I-$nosZ$II in the metatranscriptomes, respectively. We calculated Hedges

G in a meta-analysis of effect sizes for soil studies (meta::metacont), taking the lower carbon soil as the reference and carbon-enriched soil as the treatment (e.g. bulk soil or unamended soil compared to rhizosphere, glucose, glycine, or root exudate amended soils). Aquatic studies were observational, so we fit a Spearman correlation coefficient between *δnos-nir* or *δnosZ*I-*nosZ*II and environmental variables (NO_3^- + NO_2^- , phosphate, bacterial production and chlorophyll A concentrations as proxies of resource availability, and oxygen). Where communities were captured and analyzed on two different filter sizes from the same sample, we determined the correlation separately for the two size fractions but counted them as a single study in our overall effect size calculations. We used psychmeta::ma_r to calculate overall correlations in the aquatic data.

Data availability

All genomes, metagenomes, and metatranscriptomes used in this paper are publicly available in NCBI or other sources under the identifiers denoted in **Supplementary Data 2, 9** and **10**. 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). Source Data are provided with this manuscript.

Code availability

Scripts used to generate figures are available in FigShare (10.6084/m9.figshare.24922776).

Acknowledgments

This work was supported by the Swedish University of Agricultural Sciences Senior Career Grant 2019-2025 to SH and the Swedish Research Council Grant Agreements No. 2016-

03551 to SH and 2023-03627 to GP. Computing resources were provided by the Department of Forest Mycology and Plant Pathology, 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, and the Swedish National Infrastructure for Computing (SNIC) at the PDC Center for High Performance Computing, KTH Royal Institute of Technology, partially funded by the Swedish Research Council through grant agreement no. 2016-07213 (NAISS 2024/22-133).

Author contributions

Conceptualization: S.H. Data curation: G.P., A.S. and C.J. Formal analysis: G.P. Funding acquisition: S.H and G.P. 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. and S.H. Writing – review and editing: G.P., A.S., C.J. and SH.

Competing interests

The authors declare no competing interests.

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. Vitale, A. *et al.* Mapping of the Denitrification Pathway in *Burkholderia thailandensis* by Genome-Wide Mutant Profiling. *J. Bacteriol.* **202**, e00304-20 (2020).

15. 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).

16. 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.* **16**, e13221 (2024).

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. Sun, X. *et al.* Ecological dynamics explain modular denitrification in the ocean. *Proc. Natl. Acad. Sci.* **121**, e2417421121 (2024).

22. 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. *Appl. Environ. Microbiol.* **90**, e01741-23 (2023).

23. Crocker, K. *et al.* Environmentally dependent interactions shape patterns in gene content across natural microbiomes. *Nat. Microbiol.* **9**, 2022–2037 (2024).

24. Mahne, I. & Tiedje, J. M. Criteria and methodology for identifying respiratory denitrifiers. *Appl. Environ. Microbiol.* **61**, 1110–1115 (1995).

25. Pold, G. *et al.* Phylogenetics and environmental distribution of nitric oxide-forming nitrite reductases reveal their distinct functional and ecological roles. *ISME Commun.* **4**, ycae020 (2024).

26. Park, D., Kim, H. & Yoon, S. Nitrous oxide reduction by an obligate aerobic bacterium, *Gemmatimonas aurantiaca* strain T-27. *Appl Environ Microbiol* vol. 83 e00502-17 (2017).

27. 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).

28. 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).

29. Xu, Z. *et al.* Unprecedented N_2O production by nitrate-ammonifying Geobacteraceae with distinctive N_2O isotopocule signatures. *mBio* **15**, e02540-24 (2024).

30. 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).

31. 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).

32. Gralka, M., Pollak, S. & Cordero, O. X. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat. Microbiol.* **8**, 1799–1808 (2023).

33. Greening, C. *et al.* Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).

34. Aroney, S. T. N. *et al.* CoverM: read alignment statistics for metagenomics. *Bioinformatics* **41**, btaf147 (2025).

35. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

36. Nishimura, Y. & Yoshizawa, S. The OceanDNA MAG catalog contains over 50,000 prokaryotic genomes originated from various marine environments. *Sci. Data* **9**, 305 (2022).

37. 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).

38. Wan, X. S. *et al.* Particle-associated denitrification is the primary source of N_2O in oxic coastal waters. *Nat. Commun.* **14**, 8280 (2023).

39. 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).

40. 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).

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

42. Saghaï, A. *et al*. Diversity of archaea and niche preferences among putative ammonia-oxidizing Nitrososphaeria dominating across European arable soils. *Environ. Microbiol.* **24**, 341–356 (2022).

43. Wang, Z., Vishwanathan, N., Kowaliczko, S. & Ishii, S. Clarifying microbial nitrous oxide reduction under aerobic conditions: tolerant, intolerant, and sensitive. *Microbiol. Spectr.* **11**, e04709-22 (2023).

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

45. 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.* 118563 (2022).

46. Roothans, N., van Loosdrecht, M. C. M. & Laureni, M. Metabolic labour division trade-offs in denitrifying microbiomes. *ISME J.* **19**, wraf020 (2025).

47. Zhang, I. H. *et al*. Partitioning of the denitrification pathway and other nitrite metabolisms within global oxygen deficient zones. *ISME Commun.* **3**, 1–14 (2023).

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

49. 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).

50. Tiedje, J. M., Sexstone, A. J., Myrold, D. D. & Robinson, J. A. Denitrification: ecological niches, competition and survival. *Antonie van Leeuwenhoek* vol. 48 569–583 (1983).

51. Jia, M., Winkler, M. K. H. & Volcke, E. I. P. Elucidating the competition between heterotrophic denitrification and DNRA using the resource-ratio theory. *Environ. Sci. Technol.* **54**, 13953–13962 (2020).

52. Stremińska, M. A., Felgate, H., Rowley, G., Richardson, D. J. & Baggs, E. M. Nitrous oxide production in soil isolates of nitrate-ammonifying bacteria. *Environ. Microbiol. Rep.* **4**, 66–71 (2012).

53. Yoon, S., Cruz-García, C., Sanford, R., Ritalahti, K. M. & Löffler, F. E. Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO_3^-/NO_2^- reduction pathways in *Shewanella loihica*strain PV-4. *ISME J.* **9**, 1093–1104 (2015).

54. 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).

55. Saleh-Lakha, S. *et al.* Effect of nitrate and acetylene on *nirS*, *cnorB*, and *nosZ* expression and denitrification activity in *Pseudomonas mandelii*. *Appl. Environ. Microbiol.* **75**, 5082–5087 (2009).

56. Thomas, K. L., Lloyd, D. & Boddy, L. Effects of oxygen, pH and nitrate concentration on denitrification by Pseudomonas species. *FEMS Microbiol. Lett.* **118**, 181–186 (1994).

57. Wrage-Mönnig, N. *et al.* The role of nitrifier denitrification in the production of nitrous oxide revisited. *Soil Biol. Biochem.* **123**, A3–A16 (2018).

58. Venterea, R. T. Nitrite□driven nitrous oxide production under aerobic soil conditions: kinetics and biochemical controls. *Glob. Change Biol.* **13**, 1798–1809 (2007).

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

60. 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).

61. 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).

62. Wei, Q. *et al.* Molecular mechanisms through which different carbon sources affect denitrification by *Thauera linaloolentis*: Electron generation, transfer, and competition. *Environ. Int.* **170**, 107598 (2022).

63. Pan, Y., Ni, B.-J., Bond, P. L., Ye, L. & Yuan, Z. Electron competition among nitrogen oxides reduction during methanol-utilizing denitrification in wastewater treatment. *Water Res.* **47**, 3273–3281 (2013).

64. Schalk-Otte, S., Seviour, R. J., Kuenen, J. G. & Jetten, M. S. M. Nitrous oxide (N_2O) production by *Alcaligenes faecalis* during feast and famine regimes. *Water Res.* **34**, 2080–2088 (2000).

65. McGill, S. L. *et al.* Pseudomonas aeruginosa reverse diauxie is a multidimensional, optimized, resource utilization strategy. *Sci. Rep.* **11**, 1457 (2021).

66. Dufour, L. J. P. *et al.* Potential energetic return on investment positively correlated with overall soil microbial activity. *Soil Biol. Biochem.* **173**, 108800 (2022).

67. LaRowe, D. E. & Van Cappellen, P. Degradation of natural organic matter: A thermodynamic analysis. *Geochim. Cosmochim. Acta* **75**, 2030–2042 (2011).

68. Morley, N. & Baggs, E. M. Carbon and oxygen controls on N_2O and N_2 production during nitrate reduction. *Soil Biol. Biochem.* **42**, 1864–1871 (2010).

69. Curtright, A. J. & Tiemann, L. K. Chemical identity of carbon substrates drives differences in denitrification and N_2O reduction within agricultural soils. *Soil Biol. Biochem.* **184**, 109078 (2023).

70. Henry, S. *et al.* Disentangling the rhizosphere effect on nitrate reducers and denitrifiers: insight into the role of root exudates. *Environ. Microbiol.* **10**, 3082–3092 (2008).

71. Sennett, L. B. *et al.* Determining how oxygen legacy affects trajectories of soil denitrifier community dynamics and N2O emissions. *Nat. Commun.* **15**, 7298 (2024).

72. Bertagnolli, A. D., Konstantinidis, K. T. & Stewart, F. J. Non-denitrifier nitrous oxide reductases dominate marine biomes. *Environ. Microbiol. Rep.* **12**, 681–692 (2020).

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

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

75. Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

76. Sun, Y., De Vos, P. & Heylen, K. Nitrous oxide emission by the non-denitrifying, nitrate ammonifier Bacillus licheniformis. *BMC Genomics* **17**, 68 (2016).

77. Orakov, A. *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol.* **22**, 178 (2021).

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

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

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

81. 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).

82. Qin, W. *et al*. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *ISME J.* **14**, 2595–2609 (2020).

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

84. Quéméneur, M. *et al.* Diversity surveys and evolutionary relationships of *aoxB* genes in aerobic arsenite-oxidizing bacteria. *Appl. Environ. Microbiol.* **74**, 4567–4573 (2008).

85. Garber, A. I. *et al.* FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front. Microbiol.* **11**, (2020).

86. Molenda, O. *et al*. Insights into origins and function of the unexplored majority of the reductive dehalogenase gene family as a result of genome assembly and ortholog group classification. *Environ. Sci. Process. Impacts* **22**, 663–678 (2020).

87. Fisher, K. *et al*. Efficient NADPH-dependent dehalogenation afforded by a self-sufficient reductive dehalogenase. *J. Biol. Chem.* **299**, 105086 (2023).

88. Tanabe, T. S. & Dahl, C. HMS-S-S: A tool for the identification of sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Mol. Ecol. Resour.* **22**, 2758–2774 (2022).

89. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).

90. Wells, M., Kim, M., Akob, D. M., Basu, P. & Stolz, J. F. Impact of the dimethyl sulfoxide reductase superfamily on the evolution of biogeochemical cycles. *Microbiol. Spectr.* **11**, e04145-22 (2023).

91. Ospino, M. C., Kojima, H. & Fukui, M. Arsenite oxidation by a newly isolated Betaproteobacterium possessing *arx* genes and diversity of the *arx* gene cluster in bacterial genomes. *Front. Microbiol.* **10**, (2019).

92. Price, M. N., Deutschbauer, A. M. & Arkin, A. P. Filling gaps in bacterial catabolic pathways with computation and high-throughput genetics. *PLOS Genet.* **18**, e1010156 (2022).

93. Alballa, M. & Butler, G. TooT-T: discrimination of transport proteins from non-transport proteins. *BMC Bioinformatics* **21**, 25 (2020).

94. Kim, G. B., Gao, Y., Palsson, B. O. & Lee, S. Y. DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proc. Natl. Acad. Sci.* **118**, e2021171118 (2021).

95. Barnum, T. P. *et al.* Predicting microbial growth conditions from amino acid composition. 2024.03.22.586313 Preprint at https://doi.org/10.1101/2024.03.22.586313 (2024).

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

97. Brooks, M. *et al.* glmmTMB: generalized linear mixed models using template model builder. (2025).

98. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

99. Parks, D. H. *et al*. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).

100. Coleman, G. A. *et al*. A rooted phylogeny resolves early bacterial evolution. *Science* **372**, eabe0511 (2021).

101. Louca, S. castor: Efficient Phylogenetics on Large Trees. (2024).

102. Blomberg, S. P., Garland JR., T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).

103. Zeileis, A., Köll, S. & Graham, N. Various versatile variances: an object-oriented implementation of clustered covariances in R. *J. Stat. Softw.* **95**, 1–36 (2020).

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

105. National Ecological Observatory Network (NEON). Soil microbe metagenome sequences. NEON https://doi.org/10.48443/fzzj-g053 (2021).

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

107. Wilhelm, R. C. *et al.* A metagenomic survey of forest soil microbial communities more than a decade after timber harvesting. *Sci. Data* **4**, 1–8 (2017).

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

109. Brown, M. V. *et al.* Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative. *Sci. Data* **5**, 1–10 (2018).

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

111. Satinsky, B. M. *et al.* The Amazon continuum dataset: quantitative metagenomic and metatranscriptomic inventories of the Amazon River plume, June 2010. *Microbiome* **2**, 1–7 (2014).

112. Biller, S. J. *et al.* Marine microbial metagenomes sampled across space and time. *Sci. Data* **5**, 1–7 (2018).

113. 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).

114. 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).

115. 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**, 1–7 (2020).

116. 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).

117. Xu, J. *et al.* The structure and function of the global citrus rhizosphere microbiome. *Nat. Commun.* **9**, 1–10 (2018).

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

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

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

121. Bivand, R. S., Pebesma, E. & Gomez-Rubio, V. *Applied Spatial Data Analysis with R, Second Edition*. (Springer, NY, 2013).
122. Bivand, R. *et al.* rgdal: bindings for the ‘geospatial’ data abstraction library. (2022).
123. Bivand, R. & Rundel, C. *Rgeos: Interface to Geometry Engine - Open Source ('GEOS’)*. (2020).
124. 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).
125. Czech, L., Barbera, P. & Stamatakis, A. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**, 3263–3265 (2020).
126. Woodcroft, B. J., Boyd, J. A. & Tyson, G. W. OrfM: a fast open reading frame predictor for metagenomic data. *Bioinformatics* **32**, 2702–2703 (2016).
127. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. VSURF: variable selection using random forests. (2022).
128. Breiman, L., Cutler, A., Liaw, A. & Wiener, M. randomForest: Breiman and Cutler’s random forests for classification and regression. (2022).
129. Molnar, C., Bischl, B. & Casalicchio, G. iml: an R package for interpretable machine learning. *JOSS* **3**, 786 (2018).

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 covering a broad range of environmental, host-associated, and engineered biomes were extracted from GTDB v.214.1 [99] (**Supplementary Data 2**). After quality filtering the genomes, denitrification genes were searched using a combination of HMM-based searches, phylogeny construction, and manual inspection of alignments. Denitrifier genomes were then searched for additional functional traits. Maximum growth rate was inferred based on codon use bias in highly expressed ribosomal proteins, while transcription factors were detected using DeepTFactor, and potential to use various organic and inorganic electron donors and acceptors with a range of tools. **b** Venn diagram of prevalence of denitrifier types. The number inside each circle denotes the percentage of denitrifier genomes with that genotype. Letters in italics denote whether genotype is considered an "initiator" (*i*), "terminator" (*t*) or both (*it*). Reactions completed by each enzyme and corresponding denitrifier type name are listed below. Genomes encoding *nor* without accompanying *nir* or *nosZ* genes are marked with an asterisk and classified as belonging to non-denitrifiers. **c** Stacked bar charts 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 (see methods). The number of genomes corresponding to each denitrifier type is indicated above the corresponding bar. **d** Proportion of genomes within dominant phyla encoding each denitrification enzyme gene, among denitrifiers. Numbers to the right in panel d indicate the number of denitrifier genomes belonging to each phylum denoted in graph. Phyla represented

by fewer than 100 denitrifier genomes are aggregated under "other" and can be found in **Supplementary Data 11**.

Fig. 2: Genome-inferred traits of complete and partial denitrifier bacteria. a Estimated maximum growth rate of organisms based on codon usage bias. Stacked bar charts show the maximum predicted growth rate for genomes separated by denitrifier type (upper panel) and NosZ clade (lower panel). 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. Numbers to the right denote number of genomes used in maximum growth rate predictions and exclude genomes with fewer than 10 ribosomal proteins annotated. **b** Boxplots of sugar and organic acid preference of denitrifiers, based on KEGG annotation ratios[32]. **c** Genomic density of transcription factors inferred using DeepTFactor. **d** Boxplot of substrate use counts by denitrifier type and substrate category based on GapMind, with colors following panels b-c. **e** Heatmap of terminal electron acceptors and donors, with intensity of color and number denoting proportion of genomes within denitrifier type encoding redox ability. In panels b-d, box boundaries represent first and third quartiles, with midline denoting median. Whiskers denote the 1.5 IQR, and outliers are shown as individual points. Asterisks above boxes indicate a difference between complete denitrifiers and the partial denitrifier type and no comparisons between partial denitrifier types were made. Numbers to the right in panel e denote number of genomes used in panels b-e.

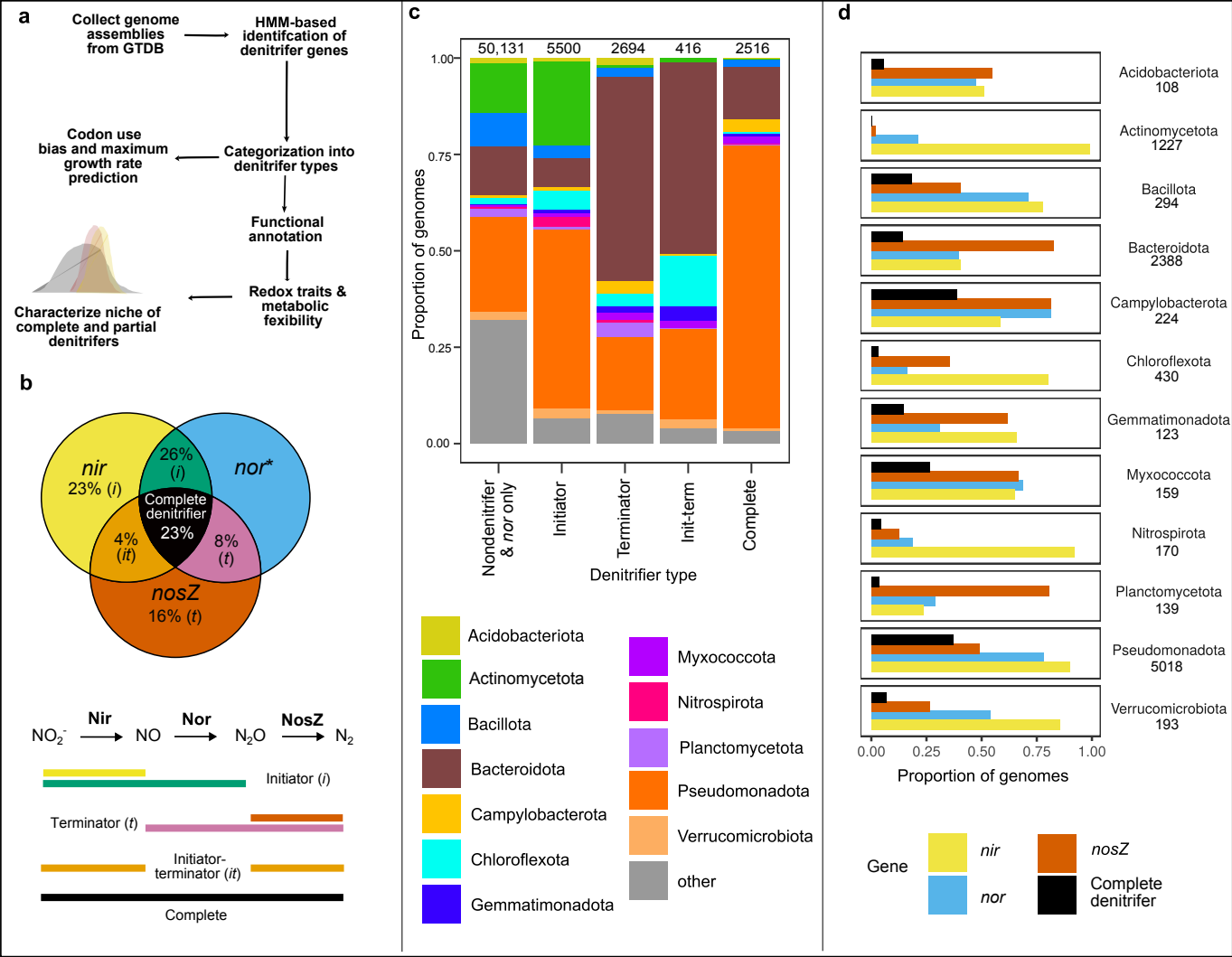
Fig. 3: Balance in prevalence of denitrification genes shows dominance of initiators at the community level across global biomes. **a, b** Boxplots showing difference in *nos* versus *nir* counts ((*nosZ* I + *nosZ* II) – (*nirK* + *nirS*)) (**a**); and clade I and clade II *nosZ* counts (**b**) per Gb sequenced. Biomes were compared using Benjamini-Hochberg FDR-corrected

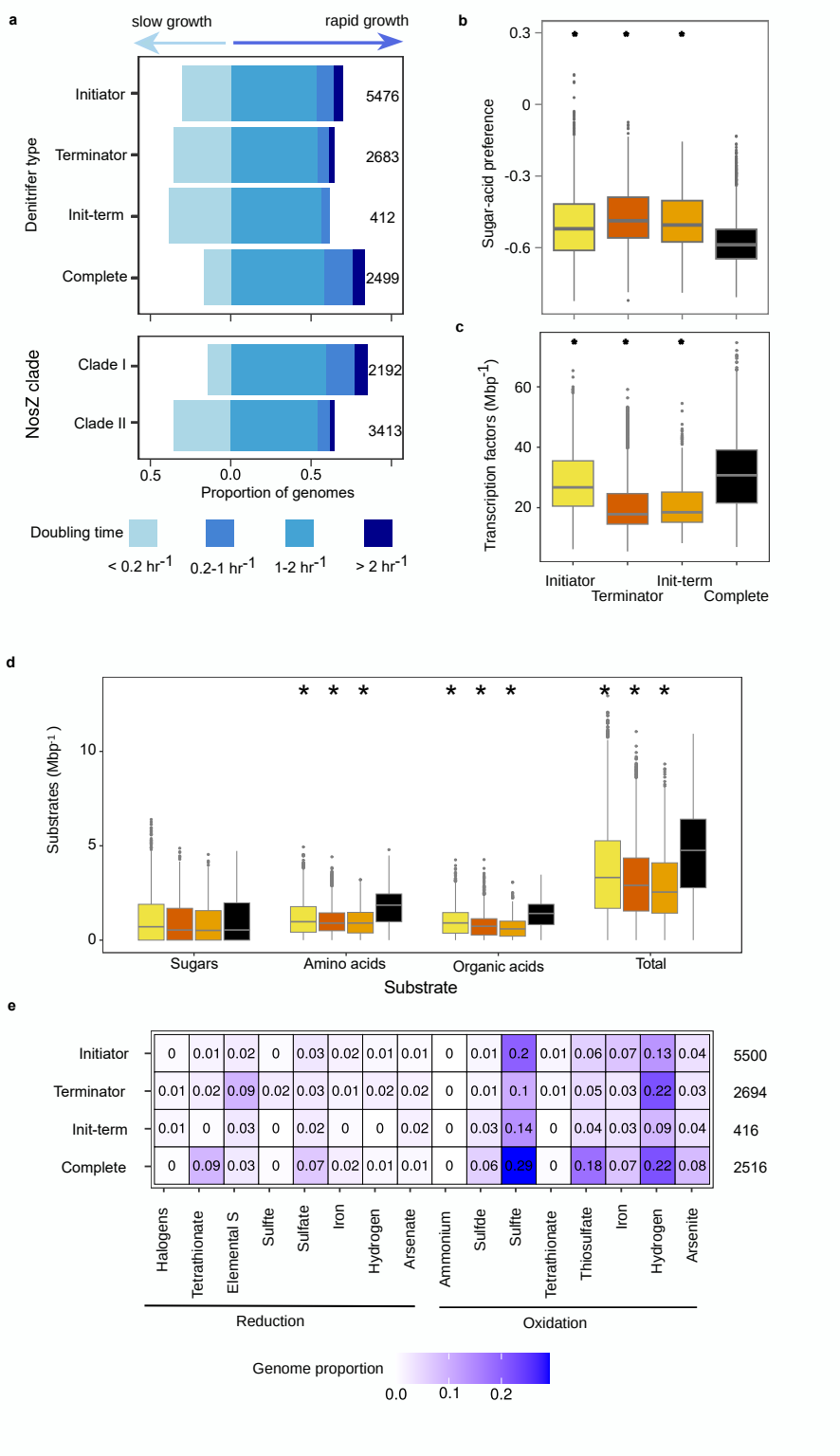
pairwise ranked comparisons following Kruskall-Wallis, and common letters to the right in each boxplot denotes biomes with similar median gene prevalences. Numbers to the right denote the number of metagenomes included for each biome. Box boundaries represent first and third quartiles, with midline denoting the median and whiskers the 1.5 IQR. **c, d** Mean phylum level composition of clade I (**c**) and clade II *nosZ* reads (**d**), 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.

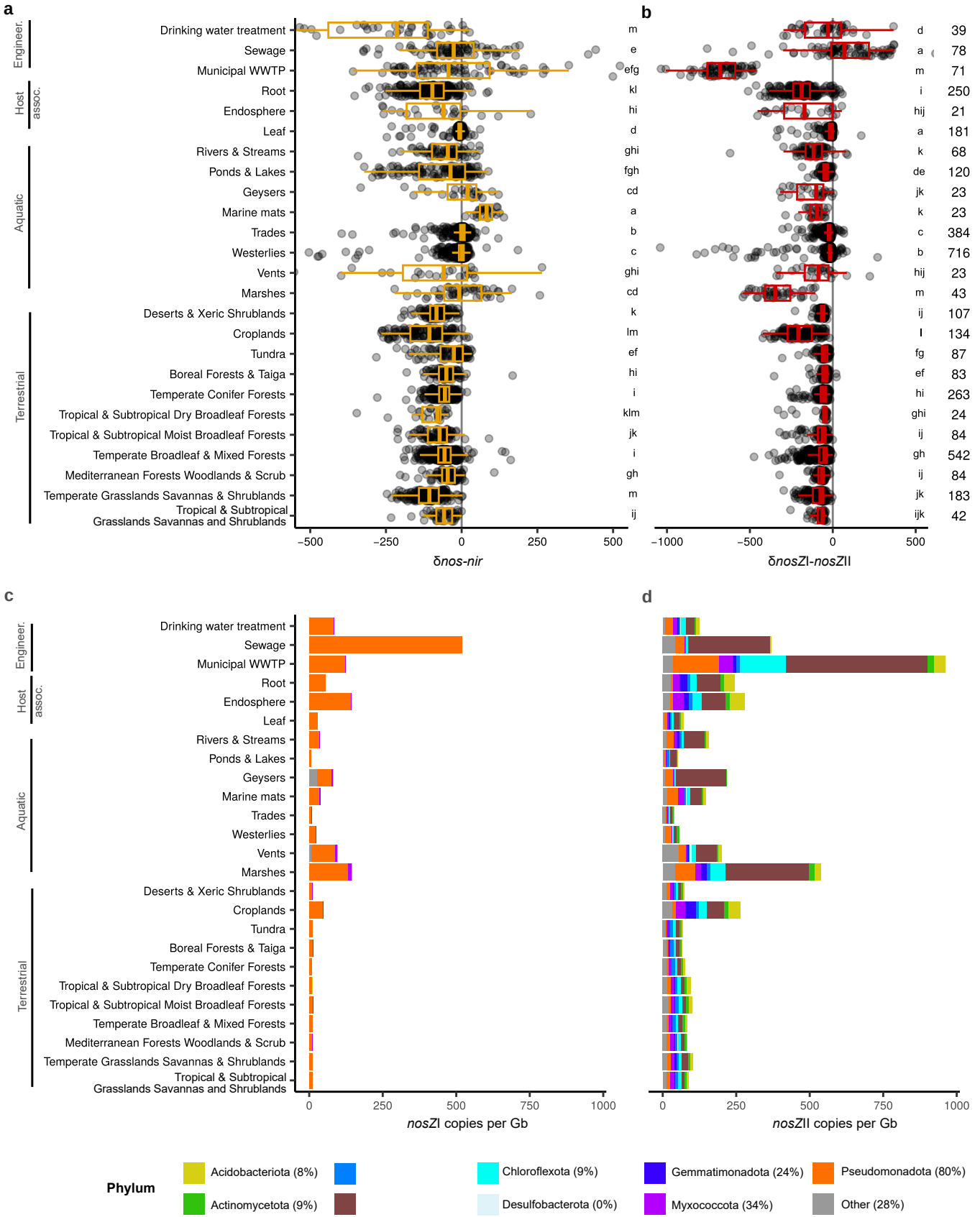
Fig. 4: Environmental predictors of balance between *nos* and *nir* counts using random forest models. a, b Abiotic predictors of the difference in *nos* and *nir* gene counts, calculated as (*nosZ* I + *nosZ* II) – (*nirK* + *nirS*) and normalized per Gb sequenced in soil (**a**) and marine metagenomes (**b**). Accumulated local effects plots 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. Vertical marks on the x axis in each panel denote data density. Five hundred trees were run, and mtry, nodesize, and sampsize parameters were 3, 10, and 176 for soil and 3,10, 88 for the marine models, respectively.

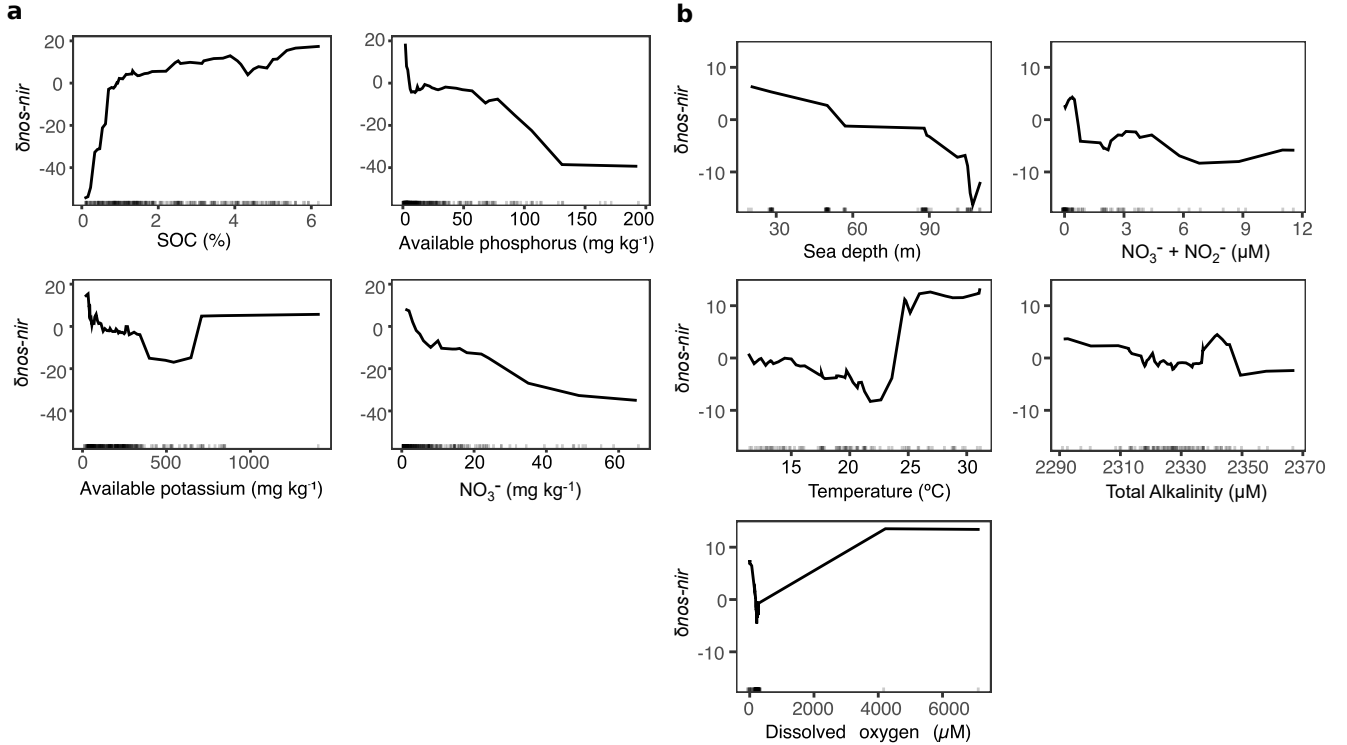
Fig. 5: Environmental predictors of balance between *nosZ* clade I and clade II using random forest models. a, b Abiotic predictors of the difference in *nosZ* clade I and II gene counts normalized per Gb sequenced in soil (**a**) and marine metagenomes (**b**). Accumulated local effects plots show the differences in prediction of the δ*nosZ*I-*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. Vertical marks on the x axis in each panel denote data density. Five hundred trees were run, and mtry, nodesize, and sampsize parameters were 4, 6 and 209 for soil and 3,6, and 88 for the marine models, respectively.

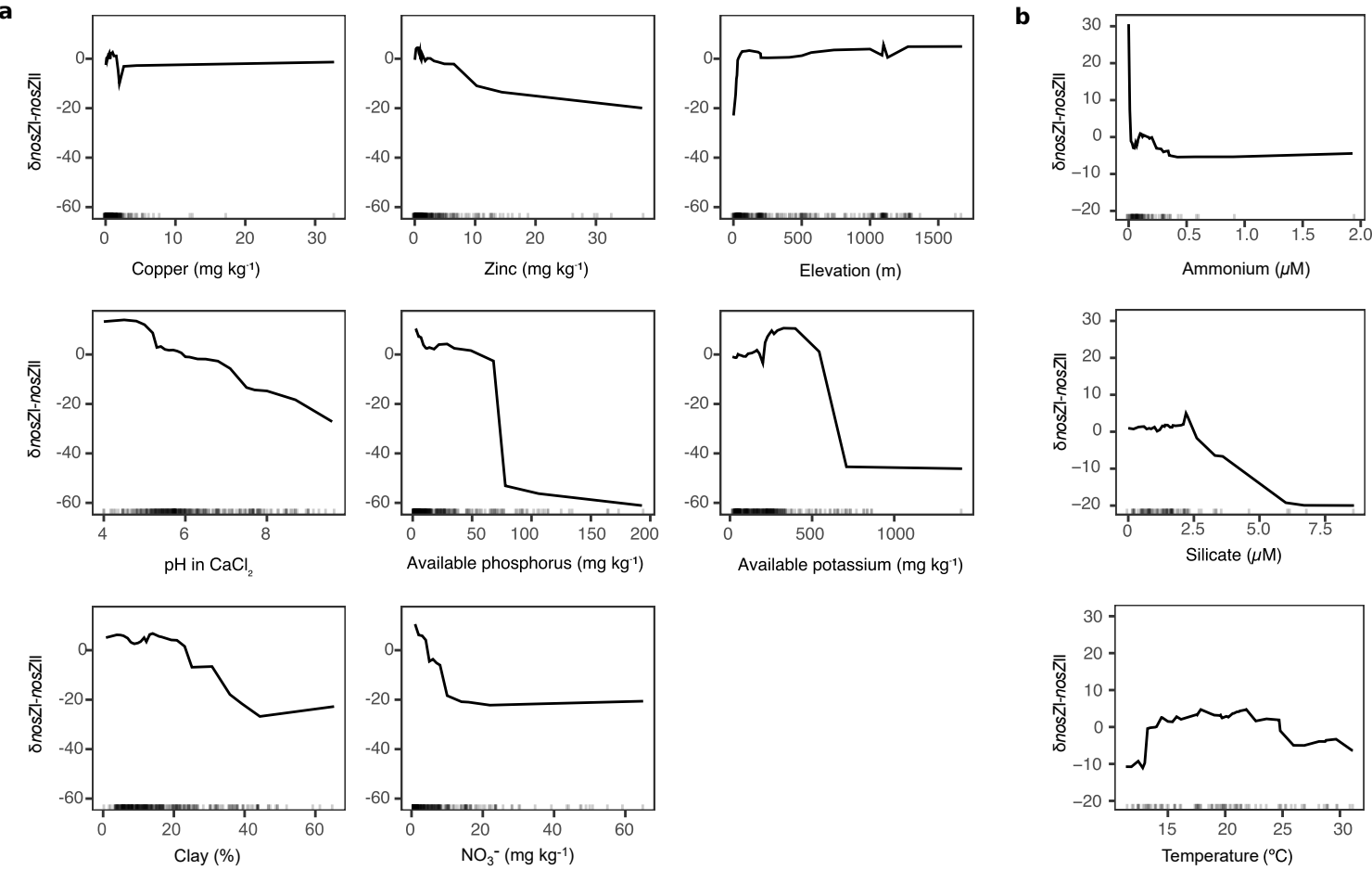
Fig. 6: Balance in prevalence of denitrification gene transcripts shows dominance of initiators at the community level in terrestrial and aquatic biomes. a δ*nos-nir* and δ*nos*ZI-*nosZ*II normalized per Gbp mRNA sequenced across different biomes. Numbers to the right denote number of metatranscriptomes included in each category. **b** Mean and 95 % confidence intervals of spearman correlation coefficients between δ*nos-nir* or δ*nos*ZI-*nosZ*II and environmental factors in aquatic metatranscriptomes. Coloring follows panel **a**. Circle size and numbers to the right denote the number of studies the effect size is calculated from. Soil samples with root exudate addition are categorized under rhizosphere in **a**. Box boundaries represent first and third quartiles, with midline denoting median and whiskers the 1.5 IQR. Values for all biomes can be found in **Supplementary Data 12**.

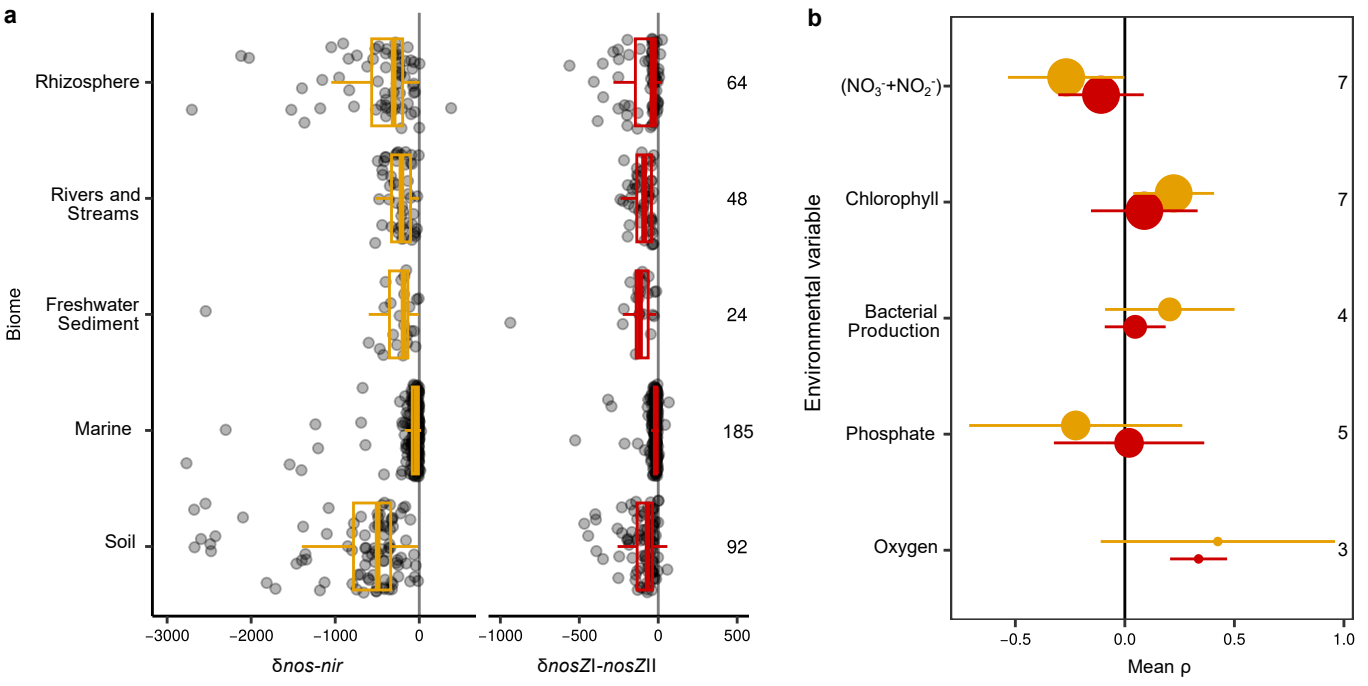

a
Collect genome assemblies from GTDB
HMM-based identifcation of denitrifer genes
Categorization into denitrifer types
Codon use bias and maximum growth rate prediction
Functional annotation
Redox traits & metabolic fexibility
Characterize niche of complete and partial denitrifers
b
nir
23% (i)
26% (i)
nor*
Complete denitrifier
23%
4% (it)
8% (t)
nosZ
16% (t)
NO2- Nir NO Nor N2O NosZ N2
Initiator (i)
Terminator (t)
Initiator-terminator (it)
Complete
c
50,131
5500
2694
416
2516
Proportion of genomes
1.00
0.75
0.50
0.25
0.00
Nondenitrifer & nor only
Initiator
Terminator
Init-term
Complete
Denitrifier type
Acidobacteriota
Actinomycetota
Bacillota
Bacteroidota
Campylobacterota
Chloroflexota
Gemmatimonadota
Myxococcota
Nitrospirota
Planctomycetota
Pseudomonadota
Verrucomicrobiota
other
d
Acidobacteriota 108
Actinomycetota 1227
Bacillota 294
Bacteroidota 2388
Campylobacterota 224
Chloroflexota 430
Gemmatimonadota 123
Myxococcota 159
Nitrospirota 170
Planctomycetota 139
Pseudomonadota 5018
Verrucomicrobiota 193
0.00
0.25
0.50
0.75
1.00
Proportion of genomes
Gene
nir
nor
nosZ
Complete denitrifer


a
slow growth
rapid growth
Denitrifer type
Initiator
Terminator
Init-term
Complete
5476
2683
412
2499
NosZ clade
Clade I
Clade II
2192
3413
0.5
0.0
0.5
Proportion of genomes
Doubling time
< 0.2 hr-1
0.2-1 hr-1
1-2 hr-1
> 2 hr-1
b
Sugar-acid preference
0.3
0
-0.3
-0.6
c
Transcription factors (Mbp-1)
60
40
20
Initiator
Terminator
Init-term
Complete
d
Substrates (Mbp-1)
10
5
0
Sugars
Amino acids
Organic acids
Total
Substrate
e
Initiator
Terminator
Init-term
Complete
5500
2694
416
2516
Halogens
Tetrathionate
Elemental S
Sulfite
Sulfate
Iron
Hydrogen
Arsenate
Ammonium
Sulfide
Sulfite
Tetrathionate
Thiosulfate
Iron
Hydrogen
Arsenite
Reduction
Oxidation
Genome proportion
0.0
0.1
0.2


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
m e efg kl hi d ghi fgh cd a b c ghi cd k lm ef hi i klm jk i gh m ij
d a m i hij a k de jk k c b hij m ij l fg ef hi ghi ij gh ij jk ijk
39 78 71 250 21 181 68 120 23 23 384 716 23 43 107 134 87 83 263 24 84 542 84 183 42
−500 −250 0 250 500
δnos-nir
−1000 −500 0 500
δnosZI-nosZII
nosZI copies per Gb
nosZII copies per Gb
0 250 500 750 1000
Phylum
Acidobacteriota (8%)
Actinomycetota (9%)
Chloroflexota (9%)
Desulfobacterota (0%)
Gemmatimonadota (24%)
Myxococcota (34%)
Pseudomonadota (80%)
Other (28%)


a
b
δnos-nir
SOC (%)
Available phosphorus (mg kg-1)
Available potassium (mg kg-1)
NO3- (mg kg-1)
Sea depth (m)
NO3- + NO2- (μM)
Temperature (°C)
Total Alkalinity (μM)
Dissolved oxygen (μM)


a
b
δnosZI-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)
Clay (%)
NO3- (mg kg-1)
Temperature (°C)


a
Biome
Rhizosphere
Rivers and Streams
Freshwater Sediment
Marine
Soil
64
48
24
185
92
−3000
−2000
−1000
0
δnos-nir
−1000
−500
0
500
δnosZI-nosZII
b
Environmental variable
(NO3-+NO2-)
Chlorophyll
Bacterial Production
Phosphate
Oxygen
7
7
4
5
3
−0.5
0.0
0.5
1.0
Mean ρ