



Published in final edited form as:

Nat Microbiol. 2024 August ; 9(8): 2022–2037. doi:10.1038/s41564-024-01752-4.

Environmentally-dependent interactions shape patterns in gene content across natural microbiomes

Kyle Crocker^{1,2,3}, Kiseok Keith Lee^{1,2,3}, Milena Chakraverti-Wuerthwein^{2,3,4}, Zeqian Li^{1,2,5}, Mikhail Tikhonov⁶, Madhav Mani^{7,8,9,10}, Karna Gowda^{11,12,*}, Sepp Kuehn^{1,2,3,10,*}

¹Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA

²Center for the Physics of Evolving Systems, The University of Chicago, Chicago, IL 60637, USA

³Center for Living Systems, The University of Chicago, Chicago, IL 60637, USA

⁴Biophysical Sciences Graduate Program, The University of Chicago, Chicago, IL 60637, USA

⁵Department of Physics, The University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁶Department of Physics, Washington University in St. Louis, St. Louis, MO 63130, USA

⁷Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA

⁸Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

⁹NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL 60208, USA

¹⁰National Institute for Theory and Mathematics in Biology, Northwestern University and The University of Chicago, Chicago, IL, USA

¹¹Department of Microbiology, The Ohio State University, Columbus, OH 43210, USA

¹²Biophysics Graduate Program, The Ohio State University, Columbus, OH 43210, USA

Abstract

Sequencing surveys of microbial communities in hosts, oceans, and soils have revealed ubiquitous patterns linking community composition to environmental conditions. While metabolic

*Correspondence: karna.gowda@gmail.com, sepp.kuehn@gmail.com.

Author contributions: K.C., K.G., M.M. and S.K. conceptualized the research. K.C., K.G., and S.K. designed the experiments. K.C., K.G., K.L., and M.C-W. performed the experiments. K.G. performed statistical analysis of the topsoil microbiome data and performed and analyzed soil enrichment experiments. M.C-W. performed mortality rate experiments, advised by K.C. and S.K.. Z.L. performed co-evolution analysis, advised by K.G. and S.K.. K.C. performed all other experiments and analyses, advised by K.G. and S.K.. K.C. performed CRM fits, predictions, and simulations, adapting code written by K.G.. K.L. sampled topsoils and performed the soil microcosm experiments. M.T. performed the analysis of sequencing data for the soil microcosm experiment in Fig. S3. The topsoil microcosm experiment was, in part, a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture. K.C., K.G., Z.L., K.L., M.T., and S.K. wrote the manuscript with input from all authors.

Data and materials availability: Data associated with this manuscript are publicly available at doi.org/10.17605/OSF.IO/N4J6F and by request. Raw sequence reads for soil enrichment experiments are deposited under NCBI BioProject ID PRJNA976277, and raw sequence reads for co-culture experiments are deposited under NCBI BioProject ID PRJNA1109838. Bacterial isolates are available by request. Gene abundance tables and environmental variables for the global topsoil microbiome dataset were provided by Bahram *et al.* (4). The SILVA rRNA database (Refs. (83–85); <https://www.arb-silva.de>) was used for taxonomic classification.

Code availability: Code associated with this manuscript is publicly available at doi.org/10.17605/OSF.IO/N4J6F and by request.

Competing interests: The authors declare no competing interests.

capabilities restrict the environments suitable for growth, the influence of ecological interactions on patterns observed in natural microbiomes remains uncertain. Here, we use denitrification as a model system to demonstrate how metagenomic patterns in soil microbiomes can emerge from pH-dependent interactions. In an analysis of a global soil sequencing survey, we find that the abundances of two genotypes trade-off with pH; *nar* gene abundances increase while *nap* abundances decrease with declining pH. We then show that in acidic conditions strains possessing *nar* fail to grow in isolation but are enriched in the community due to an ecological interaction with *nap* genotypes. Our study provides a roadmap for dissecting how associations between environmental variables and gene abundances arise from environmentally-modulated community interactions.

Introduction

Natural microbiomes drive the global cycling of carbon, nitrogen, and other elements essential to life on Earth (1). This activity arises from the collective action of metabolic pathways carried by diverse, interacting microbes in dynamic communities. Despite this complexity, global sequencing surveys of natural communities have revealed ubiquitous patterns relating the abundance of the genes that make up these pathways to local environmental variables in marine systems (2), soils (3-5) and human hosts (6-8). Uncovering the origins of environmentally-mediated variation in microbiome gene content is a necessity to understand how human activity impacts global nutrient cycles.

Conventional wisdom for the distribution of microbes is that “everything is everywhere but the environment selects” (9) – but how does the environment select? One view is that microbes occupy environmental niches that are a function of their metabolic capabilities and traits. For example, the relative abundance of acid-tolerant Acidobacteria in soil increases as the pH declines (10-12). In such cases, we presume that specific metabolic strategies are enriched in niches where those strategies facilitate an individual competitive advantage. However, we also know that microbial communities are complex systems with numerous interactions between constituents (13-16), resulting in feedbacks, counter-intuitive inhibitory effects, coevolution, and predation, all of which conspire to determine abundances (17-21). Thus environmental factors should in principle modulate interactions and drive impacts on community diversity, gene content, and metabolic activity.

From this vantage, a problem arises: how can simple patterns in environmentally mediated gene content emerge given the apparent complexity of community interactions? One possibility is that interactions are weak, and changes in gene content reflect an adaptation of individual genotypes to local environmental conditions. A second possibility is that the patterns in gene content emerge from ecological interactions whose strength and specificity are modulated by local environmental variables. In this case, the emergence of reproducible patterns requires that the interactions exhibit regularity across different locations with similar environmental conditions. However, given the complexity of interactions in microbiomes, imagining how this regularity could manifest is a challenge.

Here, we present evidence that patterns in the metagenomic content of microbial communities can emerge from environmentally dependent interactions between members

of the community. Using bacterial denitrification as a model process, we find that soil pH strongly associates with variation in denitrification reductase gene content across the global topsoil microbiome (4). Through enrichments and quantitative phenotyping of isolates, we then demonstrate how pH shapes ecological interactions and patterns in denitrification gene abundances. We argue that these interactions yield reproducible metagenome-environment associations as a result of the conserved phenotypic properties of the genotypes involved, and the interactions these physiological traits support. Our study provides a unified approach for uncovering how ubiquitous patterns in microbiomes can emerge from environmentally-modulated interactions between members of the community.

Results

Quantifying patterns in topsoil denitrification gene content

We first sought to quantify metabolic gene content variation across environments, focusing on the widespread process of denitrification in soils (22). Denitrification involves the anaerobic respiration of oxidized nitrogen compounds through a cascade of reduction reactions (Fig. 1A). Denitrification is essential to global nitrogen cycling (23), representing the primary sink of bioavailable nitrogen from the biosphere. This process also influences human health through activity in wastewater treatment (24) and the human gut (25). The fact that denitrification arises from enzymes that are well-studied and reliably annotated (26, 27), and is performed by diverse, culturable taxa (28-31), enabled us to readily identify genomic patterns in gene content from sequencing data and dissect interactions using wild isolates.

To study denitrification gene content in soils we utilized the survey by Bahram *et al.* (4) of the global topsoil microbiome. The dataset comprises $n = 189$ samples taken from sites across the planet (Fig. 1A; Ref. 4). Crucially, this dataset combines shotgun metagenomics with a detailed characterization of the physicochemical properties of each soil sample, including pH, ion concentrations, nutrient levels, and site climate characteristics (Methods). The fact that these measurements were performed in a standardized fashion on soils from across the globe enabled us to reliably relate environmental factors to the abundances of denitrification genes.

First, we computed the relative abundance of the six main reductases in the denitrification pathway (*narG*, *napA*, *nirS*, *nirK*, *norB*, and *nosZ*; Fig. 1B). Note that, due to the sequencing depth of the dataset, we were not able to construct reliable metagenome-assembled genomes, so the genomic context of these genes is not known. We refer to the total prevalence denitrification reductases in a given sample as the “magnitude” (d) of the pathway, and the relative contribution of each reductase to the total as the “composition” ($\vec{c}$, Fig. 1C). We observed that denitrification pathway magnitude varies by about 50-fold across soil samples. While pathway composition also varies from sample to sample, this variation is challenging to visually distinguish from variation in pathway magnitude.

To disentangle pathway magnitude from composition, we developed a new approach based on unit-invariant singular value decomposition (uiSVD, Ref. 32; Methods). We used this approach to decompose the relative abundances of denitrification genes in each sample ($\vec{x}_i$) into a contribution arising from the magnitude of the pathway (d_i) and the pathway

composition ($\vec{c}_i$) (Fig. 1D). We note that it was not necessary to transform the data to account for compositionality, as the high diversity of the gene abundance dataset makes the effects of compositionality negligible (Supplementary Information Section 1).

pH correlates with pathway composition on a global scale

To assess how environmental variation impacts the prevalence of denitrification reductases, we computed the principal components (PCs) of the composition matrix (Methods, Extended Data Fig. 1), which represent covarying groups of genes. Correlations between PC scores (projections of composition onto the PCs) with the environmental variables show that a single component of variation, PC2 (21.7% variance explained), was associated with pH much more strongly than any other combination of components and environmental variables ($\rho = 0.64$, $p < 10^{-6}$ via one-tailed randomization test; Fig. 1E, G). The loadings of each reductase in PC2 (Fig. 1H) reveal how denitrification reductase genes vary with pH on a global scale: some reductase genes are enriched as pH decreases (*narG*, *nirS*, *norB*), while others decline (*napA*, *nirK*, *nosZ*). While pH is widely appreciated as a strong contributor to community diversity in soils (33), and is known to be important for denitrification (34, 35), our analysis provides a granular view into how pH impacts the composition of the denitrification pathway across environments.

In contrast to the composition ($\vec{c}_i$), pathway magnitude (d_i) was most strongly associated with C/N ratio ($\rho^2 = 0.35$ Fig. 1F). The correlation is negative ($\rho = -0.59$, $p < 10^{-6}$ via one-tailed randomization test; Extended Data Fig. 2), consistent with previous work showing that the prevalence of denitrification is enhanced at low C/N ratios (36). The fact that pathway magnitude and composition are strongly associated with distinct environmental factors highlights the need to disentangle these two components to assess the drivers of their variation.

Enrichment cultures reproduce topsoil microbiome patterns

If pH is causally related to patterns of denitrification gene abundances in the global topsoil microbiome (Fig. 1E, G, H), we reasoned that enrichment cultures with different pH conditions should reproduce these patterns. We extracted microbial communities from six soil samples (Table S1), and inoculated them into chemically defined media buffered at pH 6.0 and 7.3. We then passaged these communities serially for 12 growth cycles under anaerobic conditions for 72 h each, with 8-fold dilutions between cycles, supplying 1.75 mM nitrate at the beginning of each cycle (Fig. 2A; Methods). We chose these conditions because pH in soils is often strongly buffered (37-39), mM concentrations of nitrate are often present (40,41), and the growth medium was designed to capture a diverse range of denitrifiers (29). The resulting enrichments were then sequenced via shotgun metagenomics to assay taxonomic and functional composition (Methods).

Endpoint enrichments were dominated by the orders Pseudomonadales and Rhizobiales, with the former comprising the majority of all enrichments at pH 6.0 and the latter making up the majority of five out of six enrichments at pH 7.3 (Fig. 2B). Because the starting points of enrichments for both pH conditions were the same, this result suggests that pH 6.0 selects for Pseudomonadales strains, while pH 7.3 selects for Rhizobiales strains. Taxa from

the orders Burkholderiales and Enterobacteriales were present at comparably lower levels, and no other taxonomic orders were present at a relative abundance greater than 1%.

Metagenome-assembled genomes (MAGs) (42) revealed that taxa from different enrichments had largely similar denitrification genotypes (Fig. 2C); Pseudomonadales MAGs typically encoded the reductases NarG and NirS, while Rhizobiales MAGs encoded the NapA and NirK reductases. A second experiment performed with seven soil samples enriched across a broader range of pH values also reproduced the same basic pattern in genotype abundances, with taxa possessing NarG heavily enriched at pH 5.0 and 5.5 (Extended Data Fig. 3; Methods).

Strikingly, these enrichments reproduce the patterns in denitrification gene content observed in the topsoil microbiome (Fig. 1H); namely, we observed a greater prevalence of *narG* and *nirS* at acidic pH (Pseudomonadales in Figs. 2B and C), and alternatively a greater prevalence of *napA* and *nirK* at neutral pH (Rhizobiales in Figs. 2B and C). This agreement between the enrichment and the global topsoil microbiome suggests that the mechanisms shaping these patterns in soils are also at play *in vitro*.

Individual traits do not explain acidic enrichment outcomes

To uncover the mechanisms that give rise to selection for specific genotypes in our enrichment experiments (Fig. 2C), we isolated taxa from the enrichment endpoint from the genera *Pseudomonas* (*narG/nirS*) and *Rhizobium* (*napA/nirK*) (Fig. 2D, Methods). We denote the taxonomy and genotype of these isolates as PD Nar⁺ and RH Nap⁺. The simplest interpretation for the dominance of these strains is that the Nar⁺ strain is better adapted than the Nap⁺ strain to acidic conditions, and vice versa in neutral conditions. To test this hypothesis, we performed a serial dilution experiment with isolates cultured axenically under anaerobic conditions (Methods). We measured the dynamics of nitrate, nitrite, and total biomass at the end of each cycle (Fig. 3).

Indeed, the isolate RH Nap⁺, representative of taxa most abundant in pH 7.3 enrichments, appeared better adapted for pH 7.3 than PD Nar⁺. RH Nap⁺ produced more biomass in the final cycle at pH 7.3 (Fig. 3C) and accumulated less nitrite in the first cycle (Fig. 3G), despite reducing nitrate more quickly (Fig. 3E). Additionally, simulated competition using a consumer-resource model previously developed for denitrification (31) confirmed the fitness RH Nap⁺ over PD Nar⁺ at pH 7.3 based on monoculture phenotypes (Supplementary Information Section 2.1; Extended Data Fig. 4).

In contrast, at pH 6.0 PD Nar⁺ produced only a small amount of biomass, with levels decreasing with each cycle (Fig. 3B). This strain also reduced nitrate more slowly than RH Nap⁺ in cycles 2–4, and reduced nitrite more slowly in all cycles (Fig. 3D, F). Moreover, PD Nar⁺ transiently accumulated nitrite, with accumulation increasing over successive cycles (Fig. 3F).

These experiments reveal a paradox at pH 6.0: PD Nar⁺ dominated in a community context (Fig. 2B) and yet performed poorly by itself. This result suggested that the dominance

of Nar⁺ genotypes in acidic enrichment experiments was not simply a consequence of individual traits, and instead emerged from community-level processes.

Nitrite toxicity impacts acidic pH denitrification activity

To determine how community-level processes account for the discrepancy between enrichments (Fig. 2) and monocultures (Fig. 3) at pH 6.0, we investigated mechanisms underlying the poor performance of PD Nar⁺. Given previous literature indicating the toxicity of nitrite under acidic conditions (43-45), we hypothesized that nitrite accumulation (Fig. 3F) inhibited the growth of PD Nar⁺ at pH 6.0.

We first tested whether nitrite inhibited either isolate at pH 6.0. To do this, we initiated monocultures of each strain with 1.75 mM nitrate and varying concentrations of nitrite and then measured nitrate and nitrite dynamics over 72 h of growth. For both strains, we found that inhibition increased with initial nitrite concentration (Fig. 4A). Further, the nitrate reduction rate of PD Nar⁺ slowed gradually with increasing initial nitrite concentration, and the nitrate reduction rate for RH Nap⁺ showing a threshold-like dependence on nitrite concentration ($[\text{NO}_2^-] \sim 0.4\text{-}0.5\text{ mM}$).

Next, we measured killing curves for each strain in monoculture at pH 6.0 with no nitrate and 0, 0.875, or 1.75 mM nitrite (Fig. 4B; Methods). We observed the mortality of PD Nar⁺ increased with the concentration of nitrite, while the mortality rate of RH Nap⁺ changed little. This suggests that the accumulation of nitrite by PD Nar⁺ in monoculture (Fig. 3F) causes cell death.

Finally, we performed PD Nar⁺ monocultures at pH 6.0 with either high (1.75 mM) or low (0.875 mM) concentrations of nitrate. PD Nar⁺ accumulated more nitrite in the high nitrate condition (Fig. 4C), and produced one-fifth of the biomass yield per unit substrate consumed compared with the low nitrate condition. We additionally inferred that toxicity slowed the rate of nitrite reduction using a consumer-resource model (Fig. 4C; Methods; Supplementary Information Section 2.2). Overall, these experiments demonstrated that nitrite accumulation has a significant deleterious effect on the growth and metabolic activity of PD Nar⁺.

Co-culture alleviates nitrite toxicity in acidic conditions

PD Nar⁺ grew poorly in monoculture at pH 6.0 (Fig. 3) due to toxicity from accumulating nitrite (Fig. 4), but still managed to dominate acidic enrichments (Fig. 2). We hypothesized that because RH Nap⁺ diminishes nitrite accumulation in monoculture (Fig. 4A, Fig. 3F), it does so as well in co-culture, alleviating toxicity and allowing PD Nar⁺ to dominate.

To test this hypothesis, we performed 1:1 co-cultures of PD Nar⁺ and RH Nap⁺ at pH 6.0, supplying 1.75 mM nitrate (Methods). These co-cultures accumulated substantially less nitrite relative to the monoculture of PD Nar⁺, and produced significantly more biomass (Fig. 5B). Transient nitrite levels did not exceed $\sim 0.5\text{ mM}$, which is at or below the toxicity threshold identified for RH Nap⁺ monocultures in earlier experiments (Fig. 4A). We conclude that PD Nar⁺ experienced substantially less toxicity due to nitrite in co-culture than in monoculture.

We then investigated whether diminished nitrite toxicity in co-cultures explains the dominance of PD Nar⁺ over multiple cycles of growth (Fig. 2B). To do this, we co-cultured these isolates over four cycles of serial dilution and measured relative abundances via 16S amplicon sequencing (Fig. 5D; Methods). To assess whether density-dependent effects are present, we varied initial ratios of PD Nar⁺ and RH Nap⁺, holding total initial biomass constant.

Indeed, we found that co-cultures approached a stable configuration with PD Nar⁺ dominating (Fig. 5E). After the first cycle of growth, relative abundances of PD Nar⁺ decreased across most starting conditions, an effect perhaps attributable to physiological adaptation to anaerobic conditions (46). However, all conditions subsequently approached a state with PD Nar⁺ at 60–75% relative abundance by the fourth cycle. The amount of nitrite accumulation in these experiments increased with the relative abundance of PD Nar⁺ (Extended Data Fig. 5G). Similar serial co-cultures performed at pH 7.3 showed RH Nap⁺ driving PD Nar⁺ to extinction after four cycles (Fig. 5F), corroborating our expectations from earlier experiments and analyses (Fig. 3C, E, G; Extended Data Fig. 4) that RH Nap⁺ is a better competitor for resources at neutral pH.

Nar⁺ and Nap⁺ phenotypes are conserved across diverse taxa

We demonstrated that the dominance of PD Nar⁺ under acidic conditions required the presence of RH Nap⁺ (Figs. 3-5). Underlying the interactions between these isolates was the fact that the former is a nitrate specialist, capable of quickly consuming the primary limiting resource supplied to the community, while the latter is a nitrite specialist, which mitigates the toxic effects of accumulated nitrite. To extend the applicability of this observation to diverse taxa in soil microbiomes, we asked whether other strains with Nar⁺ and Nap⁺ genotypes display nitrate and nitrite specialist phenotypes, respectively.

We phenotyped 7 strains isolated in our previous study (31) which span α , β and γ —proteobacteria and possess one or the other nitrate reductase (3 Nar⁺, 4 Nap⁺). To avoid ambiguity, we excluded strains possessing both Nar and Nap. These reductases tend to appear separately in reference genomes (Supplementary Information Section 3; Fig. S1), and when they co-occur Nap may be used for redox balancing (47) or expressed primarily in μ M nitrate concentrations (48, 49). For each strain, we measured nitrate and nitrite dynamics at pH 6.0 supplied with 1.75 mM nitrate, using the area under the curve (AUC) of these dynamics to summarize traits (Fig. 6A). As with PD Nar⁺ and RH Nap⁺, we found that Nar⁺ genotypes exhibited fast nitrate reduction and the transient accumulation of nitrite, while Nap⁺ strains exhibited slow nitrate reduction and no nitrite accumulation (Fig. 6B). We conclude that Nar⁺ genotypes tend to accumulate nitrite via fast nitrate reduction, and that Nap⁺ strains tend to avoid this accumulation.

Soil microbiomes harbor substantial taxonomic diversity, raising the question of whether the trait conservation we observe in Proteobacteria (Fig. 6) is relevant to communities in the wild. To extend the applicability of our observations, we demonstrated that Proteobacteria are likely representative of denitrifying populations in natural soils using two culture-free approaches (Supplementary Information Section 4). Briefly, we found that (a) a substantial fraction of denitrification reductase reads can be classified as Proteobacterial (Fig. S2),

and (b) laboratory incubation experiments performed on homogenized topsoils under anaerobic conditions show Protobacterial populations dominate community response to nitrate amendment (Fig. S3). Finally, a meta-analysis of culture-based studies of denitrifiers in soils also supports the claim that Proteobacteria dominate these populations (Table S2). These findings support the hypothesis that the traits of our Proteobacterial isolates are representative of the dominant denitrifying populations in soils.

Discussion

We have presented a data-driven and experimentally validated approach to understanding how environmental conditions determine community composition at the granular level of gene content. Beginning with a statistical analysis of the global topsoil microbiome, we found that pH is strongly correlated with the composition of the reductases in the denitrification pathway. We then reproduced the observed pattern between pH and gene content via enrichment experiments, allowing us to dissect community responses to pH *in vitro*. Experiments on isolates from these enrichments showed that the dominance of a nitrate specialist Nar^+ isolate in acidic conditions depends on the presence of a nitrite specialist Nap^+ isolate to alleviate pH-dependent nitrite toxicity. Finally, we provided evidence that Nar^+ and Nap^+ genotypes are generically associated with nitrate and nitrite specialist phenotypes, respectively, providing a mechanistic explanation for covariation between pH and gene content in the global topsoil microbiome.

A central finding of our study is that widespread patterns in metagenomic gene content can emerge from ecological interactions between members of the community, and not necessarily only from individual adaptations to environmental conditions (50). While individual fitness effects clearly play a role in defining patterns in many contexts, interactions are likely essential to explaining patterns in a variety of other contexts. It will be an important avenue for future work to use the methods developed here to generate concrete hypotheses for the role of individual fitness effects and interactions in forming gene content patterns across a variety of metabolic processes and biomes.

Despite the breadth of our study, however, there were limitations that led us to perform additional analyses. First, our *in vitro* investigations focus on just two taxa. However, soils support thousands of taxa, raising the question of how an interaction between two taxa can be relevant in the complex ecological context of soils. Our soil microcosm experiment provides strong evidence that the complexity of the nitrate-utilizing community in soils is low (Fig. S3; Supplementary Information Section 4). Specifically, a small fraction of all taxa in the soil community respond to the addition of nitrate ($0.3\% \pm 0.2\%$, Fig. S3C). The result suggests that despite the ecological complexity of soils, a small number of taxa (4 ± 3 in our experiment, Fig. S3C) participate significantly in this metabolic process *in situ*.

Next, we have used the enzymes Nar and Nap as markers for nitrate and nitrite specialists, respectively. While it is possible for these enzymes to co-occur, we showed via a co-evolution analysis in reference genomes that genes encoding these enzymes are anti-correlated (Fig. S1), and previous work suggests that Nap may be used for redox balancing (47) or μM nitrate concentrations (48,49) in strains that contain both enzymes. Additionally,

the nitrite reductases encoded by the genes *nirS* and *nirK* must also contribute to the nitrate and nitrite specialist phenotypes. *In vitro* studies show that NirK enzymes have higher activity than NirS (26,51-64), and previous work correlates the presence of *nirK* with faster nitrite reduction traits across diverse taxa (65). However, this work also shows that strains with *narG* exhibit faster nitrate reduction than strains with only *napA* (65), a finding corroborated by *in vitro* comparisons of these enzymes (66-70). We note that one strain considered in Fig. 6, PDM21, possesses both *narG* and *nirK*, unlike either of the enrichment isolates PD Nar⁺ (which possesses *nirS*) or RH Nap⁺ (which possess *nirK*). PDM21 showed a clear nitrate specialist phenotype, despite the fact that it possesses *nirK*. The fact that Nar⁺ strains are nitrate specialists at low pH, independently of the nitrite reductase they encode, may ultimately arise because the Nar nitrate reductase is protected from extracellular pH in the cytoplasm, while the periplasmic Nap and Nir reductases are not (26).

Finally, a key finding of our study is that the interaction between Nar⁺ and Nap⁺ genotypes arises due to specialization by these two genotypes for the first and second steps of the denitrification cascade. Previous studies have suggested that such specialization is the consequence of a physiological trade-off (35, 71). We investigated this hypothesis further using simulations that account for a trade-off between nitrate and nitrite reduction rates, as well as the toxicity of nitrite (Supplemental Information). The results of these simulations are consistent with the results of our enrichment experiments: namely, specialists are selected for when toxicity is high and generalists are selected when toxicity is low (Figs. S4, S5; Supplementary Information Section 5). An important direction for future work will be experimentally characterizing the physiological basis for phenotypic specializations in denitrification.

Methods

Analysis of the global topsoil microbiome

Data pre-processing—Gene abundance tables and environmental variables for the global topsoil microbiome dataset were provided by Bahram *et al.* (4), who sampled, sequenced, and analyzed 189 topsoil sites representing the world's terrestrial biomes. Samples were processed according to standardized protocols, including shotgun metagenomic sequencing of soil DNA extracts, and chemical analysis of soil pH, P, K, Ca, Mg, ¹²C, ¹³C, ¹⁴N, and ¹⁵N. In addition, site data including mean annual temperature (MAT), mean annual precipitation (MAP), potential evapotranspiration (PET), net primary productivity (NPP), and moisture were obtained from public databases. Paired metagenome reads were quality filtered and annotated to yield KEGG ortholog (KO) abundance tables. The details of all protocols and analyses to obtain gene abundance tables and site environmental variables are given in Ref. 4.

In the subsequent analyses, the abundance of KOs corresponding to the six reductases in the denitrification pathway (*narG*/K00370, *napA*/K02567, *nirS*/K15864, *nirK*/K00368, *norB*/K04561, *nosZ*/K00376) were considered. The relative abundances of these KOs were computed by dividing the number of reads mapped to each KO by the total number of reads in each sample. In addition, 17 site environmental variables from the measurements

and analyses described above were considered: pH, P, K, Ca, Mg, C, N, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, C/N, latitude, longitude, MAT, MAP, PET, NPP, and moisture.

Unit-invariant singular value decomposition—Unit-invariant singular value decomposition (uiSVD; Ref. 32) was applied to the denitrification reductase relative abundance matrix (X) to separate variation in pathway magnitude (i.e., variations in total reductase content) from variation in pathway composition (i.e., variations in relative fractions of reductase content). This approach was motivated by modeling the data as

$$\vec{x}_i = d_i \vec{c}_i, \quad (1)$$

where $\vec{x}_i$ is the i th row of X (with entries being relative abundances of each KO), d_i is a scalar representing pathway magnitude and $\vec{c}_i$ is a vector representing pathway composition. In matrix form, this can be written as

$$X = DC, \quad (2)$$

where D is a diagonal matrix. Expressing X in this form can be achieved by applying uiSVD, which decomposes X as

$$X = D\tilde{U}\tilde{S}\tilde{V}^T E, \quad (3)$$

where D , $\tilde{S}$, and E are diagonal matrices and $\tilde{U}$ and $\tilde{V}$ are unitary matrices. Hence

$$C = \tilde{U}\tilde{S}\tilde{V}^T E. \quad (4)$$

The results of this decomposition are shown in Fig. 1D. Note that, in this representation, D encodes the scale of the rows of X (i.e., pathway magnitude), while E encodes the relative scales of the columns (i.e., typical relative scales of each gene). Code used to perform uiSVD was adapted for Python from the MATLAB code given in (32).

Having decomposed the relative abundance data X into elements of pathway magnitude (D) and composition (C), modes of gene covariation within the composition matrix were identified via principal components analysis (PCA). Because E encodes scaling information about the genes in X , $\tilde{C} \equiv CE^{-1}$ can be interpreted as a normalized gene composition matrix, for which all genes have been set to the same approximate scale. Therefore PCA was applied to $\tilde{C}$ by meancentering the columns, and applying singular value decomposition the resulting matrix, yielding a decomposition USV^T . The columns of V represent the orthonormal principal components (PCs) of $\tilde{C}$ (shown in Extended Data Fig. 1), and

$$T = US$$

(5)

represents the “scores” matrix, i.e., the projections of $\tilde{C}$ onto the PCs in V .

Analyzing gene-environment covariation—In order to relate gene covariation to variation in site environmental variables, Pearson correlations (ρ^2) were computed between the scores of each principal component (Eq. 5) and each environmental variable (squared Pearson correlations shown in Fig. 1E). Additionally, the same analysis was performed between the diagonal elements of D and each environmental parameter (Fig. 1F).

Each correlation was computed using the $n = 189$ observations from the global topsoil microbiome dataset, with environmental variables containing NaN values removed when computing each correlation. The significance of associations between C/N ratio and D , as well as between pH and PC2 scores was determined via a one-tailed randomization test. Explicitly, a distribution for the null hypothesis of zero correlation was empirically constructed by repeatedly (10^6 times) computing correlations between shuffled versions of the variables.

Phylogenetic classification of denitrification reductases—Raw reads from the global topsoil microbiome dataset (PRJEB18701) were trimmed of adapters and low quality sequences using Trimmomatic ver. 0.39 with default settings. Trimmed reads were assembled using SPAdes ver. 3.15.0. Predicted ORFs on assembled contigs were then annotated as KEGG ortholog groups (KOs) of the denitrification reductases (*narG*/K00370, *napA*/K02567, *nirS*/K15864, *nirK*/K00368, *norB*/K04561, *nosZ*/K00376) using eggNOG-mapper ver. 2.0.8, and reads were mapped to these annotated ORFs using minimap2. Finally, reads mapping to denitrification KOs were phylogenetically classified using the Kaiju web server (<https://kaiju.binf.ku.dk/server>; Ref. 72). The results of these classifications, at the level of taxonomic phylum and grouping all samples together, are shown in Fig. S2.

Enrichment and isolation of denitrifying strains

Processing of soils for primary enrichment experiments—Six forest and prairie soil samples were collected from Meadowbrook Park, Urbana, IL. Details regarding these soil samples are given in Table S1. Soils were sampled from a depth of 1–5 cm using autoclaved steel laboratory spatulas, and then stored in sealed plastic bags at 4 °C for approximately two months prior to the start of the experiment. No sample compositing was performed. To mechanically homogenize soils prior to enrichment, 5 g samples of each soil were added to sterile 50 mL centrifuge tubes, along with 25 mL PBS (pH 7.4) and 5–10 g sterile 4 mm glass beads. Tubes were then vortexed (Vortex-Genie 2) at high speed for 1 min to homogenize the samples. After vortexing, large particles in homogenized soil samples were allowed to settle for 20 min before transferring 1 mL of supernatant to sterile microcentrifuge tubes.

Primary enrichment experiments—Wells of a sterile 96-deepwell plate (Axygen PDW20C) were loaded with 1.2 mL of defined media. The media contained succinate

(25 mM) as the carbon source and 2.0 mM sodium nitrate was initially supplied. Medium pH was buffered by phosphate (40 mM) at two conditions, pH 6.0 and 7.3. This medium will hereafter be referred to as succinate defined medium (SDM); its precise composition is described in Ref. 31, and it was developed to capture a diverse range of denitrifiers (29). Six wells of each pH condition were then inoculated with 10 μ L of each soil supernatant, with two wells of each pH condition left as no-growth controls. The plate was then sealed with a gas-permeable sterile membrane (Diversified Biotech BERM-2000). After sealing, the plate was immediately transferred to an anaerobic glove box (Coy Laboratory Products 7601-110/220), which was continuously purged by a 99%/1% N₂/CO₂ mixture. The plate was incubated at 30 °C and shaken at 950 RPM (Talboys Professional 1000MP, 3 mm orbital radius).

Cultures were grown under these conditions for 72 h. At the end of this time, 150 μ L of the cultures were passaged under anaerobic conditions into a freshly prepared plate containing 1050 μ L of fresh medium (1/8 dilution factor). These passaged cultures were sealed and returned to incubation and shaking for another 72 h growth cycle. At the end of each cycle, unused cultures were assayed for endpoint nitrate and nitrite concentrations via Griess assay and vanadium (III) chloride reduction method via the protocol described in Ref. 31 (Fig. S6). In addition, optical densities at 600 nm were also recorded (BMG CLARIOstar) using 300 μ L of endpoint cultures in 96 well optical plates (Fig. S6).

Cultures were repeatedly passaged and assayed in this manner for 12 cycles. At the endpoints of cycles 4, 8, and 12, 100 μ L samples of endpoint cultures were cryopreserved by adding 100 μ L of 50 % glycerol and storing at –80 °C. At the end of cycle 12, cultures remaining after cryopreservation and nitrate/nitrite assay were frozen at –20 °C for subsequent DNA extraction.

Sequencing and analysis of primary enrichment experiments—Frozen enrichment endpoint (cycle 12) cultures stored at –20 °C were thawed and DNA extracted using the DNeasy UltraClean Microbial Kit (Qiagen). DNA concentrations were quantified using the Qubit dsDNA BR Assay Kit (Invitrogen). Library preparation for sequencing of DNA extracts was performed using the Nextera DNA Prep Kit and the Nextera DNA CD Indexes (Illumina). Pooled libraries were then sequenced using a NextSeq 500/550 Mid Output Kit v2.5 (Illumina, 2 $\times$ 150 bp paired-end), with a 1.5 pM library loading concentration and a 1% spike-in of PhiX Control v3 (Illumina). Sequencing was performed on a locally maintained and operated Illumina NextSeq 550 system.

To characterize the taxonomic composition of enrichments, 16S rRNA fragments (miTAGs) were extracted from the sequencing data using the miTAGs extraction script ver. 1 (73), after trimming and merging of overlapping paired-end reads via Trim-Galore ver. 0.6.7 (74) and BBMerge ver. 38.22 (75), respectively. The resulting miTAGs were then taxonomically classified using the RDP Classifier (76), using a confidence threshold of 80% and the copy number adjustment option enabled.

The sequencing data were then binned into metagenome assembled genomes (MAGs) using the metaWRAP pipeline ver. 1.3 (42). In summary, the pipeline performed adapter

and quality trimming of raw reads using Trim-Galore ver. 0.6.7 (74), assembly using metaSPAdes ver. 3.15.4 (77), binning using CONCOCT ver. 1.1.0 (78), MaxBin2 ver. 2.2.7 (79), and metaBAT ver. 2.15 (80), bin quality assessment using CheckM ver. 1.0.11 (81), and bin refinement by combining the results of the three binning algorithms. Bin completion and contamination thresholds of 95% and 5%, respectively, were used to obtain high-quality MAGs in most samples (Table S3). High-quality MAGs were then annotated using the RAST server (82).

Isolation of strains from primary enrichment experiments—Strains representing the dominant taxa in the enrichment experiments were isolated from cryopreserved cycle 12 samples from soil #1. Glycerol-cryopreserved cultures were streaked to purity on 1/10x TSB agar plates (1.5% agar w/v) in aerobic conditions. Overnight cultures of single-colony isolates were grown in 1/10x TSB (30 °C, 400 RPM) in aerobic conditions and cryopreserved (50% glycerol, -80 °C). Sanger sequencing of the 16S rRNA gene using 27F and 806R universal primers was used to taxonomically classify isolates using the SILVA rRNA database (83-85). RH Nap⁺ was determined to be of the family *Rhizobiaceae* and PD Nar⁺ was determined to be of the genus *Pseudomonas*.

Additional enrichments across a broader pH range—Additional enrichments using a different set of soil samples were performed across a broader range of pH conditions, spanning pH 5.0 to 7.3. Seven soil samples (soils #7–13, Table S1) taken from prairies and forest preserves across the Midwestern United States (IL, IN, MI, WI) from a depth of 1–5 cm using autoclaved steel laboratory spatulas, and were homogenized and processed as described above. These soils were used to inoculate an enrichment experiment with four different pH-buffered SDM conditions: 5.0, 5.5, 6.0, 7.3. In pH 5.0 and 5.5 conditions, succinate/succinic acid serves as the buffering agent because the buffering capacity of phosphate is weak in this pH range. Details of incubation, passaging, sampling, cryopreservation, as well as of sequencing and data analysis of cycle 12 cultures, are as described above, except that defined media succinate concentration is reduced to 4mM. Since nitrate is the limiting resource in these experiments, this change does not impact growth, but a lower succinate concentration reduces spontaneous protonation at low pH. The taxonomic composition and median MAG genotypes of endpoint enrichments are shown in Extended Data Fig. 3.

Characterization of isolate phenotypes

Culturing protocol—Strains were pre-cultured in two stages under aerobic conditions prior to transfer to denitrifying (anaerobic) conditions for phenotyping. First, wells of a sterile 24-well plate (Thermo Scientific Nunc Non-Treated Multidishes) were loaded with 1.7 mL of R2B medium. Wells were inoculated with isolates PD Nar⁺ and RH Nap⁺ from glycerol stocks stored at -80 °C. The plates were then sealed with a gas-permeable sterile membrane (Breathe-Easier, USA Scientific, 9126-2100). After sealing, the culture was incubated overnight at 0.5 rcf (400 RPM in Fisherbrand Incubating Microplate Shakers 02-217-759, 3 mm orbital radius or 219 RPM in Heidolph Unimax 1010, 10 mm orbital radius) and 30 °C in aerobic conditions. These cultures reached saturation during this time. Second, wells of a sterile 24-well plate were loaded with 1.7mL of SDM at pH 7.3 with

25 mM succinate (and no sodium nitrate). Wells were then inoculated with 17 μ L of the saturated R2B PD Nar⁺ and RH Nap⁺. After sealing, the cultures were incubated at 0.5 rcf and 30 °C in aerobic conditions overnight. These cultures reached saturation during this time. Saturated SDM culture densities were measured and normalized to an optical density of 1.0 (measured at 600 nm) via dilution into pH 7.4 phosphate-buffered saline (8 g/L H₂O, 0.2 g/L KCl, 2.68 g/L Na₂HPO₄·7H₂O, 0.24 g/L KH₂PO₄).

Due to a lag time associated with growth in anaerobic conditions for facultative anaerobes (86-91), an additional period of pre-culture in anaerobic conditions was performed before phenotyping. Wells of a sterile 96-deepwell plate (Axygen PDW20C) were loaded with 1.2 mL SDM (4mM succinate) supplemented with 1 mM sodium nitrate, from stock that had been allowed to equilibrate in the anaerobic glovebox. SDM without nitrate was loaded into at least three wells of the plate as a control for growth on trace quantities of oxygen; since succinate is a non-fermentable carbon source, any growth in the absence of nitrate in this medium indicates aerobic growth. These wells were inoculated in the glovebox with 12 μ L of OD-normalized PD Nar⁺ and RH Nap⁺ aerobic pre-cultures, resulting in a starting OD of 0.01. Additional wells were left blank as no-growth controls. Plates were sealed with a gas-permeable sterile membrane. Cultures were incubated at 30 °C and shaken at 950 RPM (Fisherbrand Incubating Microplate Shakers 02-217-759 or Talboys Professional 1000MP, 3 mm orbital radius) for 72 h.

Anaerobic pre-cultures were then used to initiate denitrification phenotyping experiments. Optical densities of endpoint anaerobic pre-cultures were measured using 300 μ L of cultures in 96-well optical plates. Optical densities were then normalized to 0.04 via dilution into SDM (no nitrate). 150 μ L of this normalized endpoint anaerobic culture was then passaged to 1050 μ L of SDM, with pH and nitrate/nitrite concentrations varying according to the experiment (e.g., Fig. 4). In the case of co-culture phenotyping experiments, this 150 μ L inoculum contained a 1:1 ratio of PD Nar⁺ and RH Nap⁺, as measured by optical density. Inoculated plates were then incubated at 30 °C and shaken at 950 rpm in anaerobic conditions. Nitrate and nitrite concentrations were assayed over time via manual sampling and subsequent Griess assay and vanadium (III) chloride reduction via the protocol described in Ref. 31. Endpoint biomass was measured via optical density at 600 nm using 300 μ L samples.

Additional experiments were conducted where measurements were taken over multiple cycles of growth and dilution into fresh medium (e.g., Figs. 3 & 5). For these experiments, aerobic pre-cultures were used to directly inoculate the first cycle of anaerobic growth. Nitrate and nitrite measurements were taken beginning in the first cycle of growth, and normalization of optical density after the first cycle was omitted. For the first cycle, 1.2 mL of SDM (4 mM succinate, 2 mM nitrate) was loaded into a 96-deepwell plate as described above. 12 μ L of normalized culture from the endpoint of the aerobic pre-culture protocol was used to inoculate the wells of this plate. For the experiment shown in Fig. 5, the fractions of PD Nar⁺ and RH Nap⁺ were chosen to achieve specified relative abundances by optical density. Two no-growth controls that contain either no inoculum (to control for contamination) or no nitrate or nitrate (to control for aerobic growth) were included on each plate.

At the end of each growth experiment or cycle, cultures were removed from anaerobic conditions, sealed with Microseal 'B' Seals (BIO-RAD, MSB1001), and stored at -80°C for subsequent DNA extraction.

Quantification of relative abundance dynamics in co-culture—Relative abundances were quantified in co-culture experiments using 16S amplicon sequencing. Frozen endpoint co-culture enrichments stored at -80°C were thawed and DNA extracted using the DNeasy 96 Blood & Tissue Kit (Qiagen). DNA concentrations were quantified using the Qubit dsDNA BR Assay Kit (Invitrogen). Next, the 16S v4 region of the rRNA gene was amplified with 515F and 806R primers using the Illumina 16S sequencing protocol (92). Briefly, a fragment of the 16S rRNA gene was amplified using the 515F (GTGYCAGCMGCCGCGGTAA) and 806R (GGACTACNVGGGTWTCTAAT) universal primers. The following reagents were used for each reaction: 14.5 μL nuclease-free H_2O , 1 μL 515F primer (5 μM), 1 μL 806R primer (5 μM), 10.5 μL DNA extract, 12.5 μL Platinum Hot Start PCR Master Mix (Invitrogen). The following thermocycler settings were used: initial denaturation, 3 min at 95°C ; amplification (25 cycles), 30 sec at 95°C , 30 sec at 55°C , 30 sec at 72°C ; final extension, 5 min at 72°C . PCR products were cleaned using the PCR CLEAN DX Kit (Aline). 16S amplicons were barcoded for multiplexed Illumina sequencing using the Nextera XT Index kit (Illumina) following the Illumina protocol (92). Pooled amplicon libraries were then submitted to the University of Chicago Genomics Facility for sequencing using a 15-20% spike-in of PhiX Control.

Resulting paired-end reads were merged with Pear (93) and then quality filtered with DADA2 plugin in the QIIME2 pipeline with default parameters (94). Taxonomy of ASVs were assigned by q2-feature-classifier prefitted to the SILVA database using the naive Bayes algorithm for the V4 region of 16S rRNA (95). ASVs of the family *Rhizobiaceae* were classified as RH Nap⁺, ASVs of the genus *Pseudomonas* were classified as PD Nar⁺. Virtually no contamination (defined as ASVs classified as neither RH Nap⁺ nor PD Nar⁺) was observed ($< 0.2\%$ total counts, $< 0.8\%$ counts per condition).

Systematic measurement bias complicates the interpretation of 16S amplicon data as a reflection of biomass relative abundance (96). Therefore we sought to infer the relative abundances by biomass using a standard curve to infer conversion factors between ASV counts and biomass for each strain. Following Ref. 96, we assumed the biomass x_* of each strain is proportional to the ASV counts C_* , with bias factors b_* :

$$\begin{aligned} x_{\text{PD}} &= b_{\text{PD}} C_{\text{PD}} \\ x_{\text{RH}} &= b_{\text{RH}} C_{\text{RH}} \end{aligned} \quad (6)$$

The relative abundance then becomes

$$f_{\text{PD}} = \frac{C_{\text{PD}}}{C_{\text{PD}} + b_{\text{RH}} C_{\text{RH}}} \quad (7)$$

where b_r is the ratio of the bias factors $b_r = b_{RH} / b_{PD}$. Using measured counts for cultures with known PD Nar⁺ relative abundances, we fit b_r using the Levenberg-Marquardt algorithm (Fig. S7) and obtained a bias ratio of 4.3 ± 0.7 . Eq. (7) was then used to infer biomass relative abundances from ASV counts. Error in inferred relative abundance comes from (a) error in the inference of b_r , and (b) variance between replicates. To account for (a), error in b_r was inferred from the covariance of the fit, and corresponding uncertainty in relative abundance was calculated using the variance formula (97). To account for (b), standard deviation of relative abundance between biological replicates was calculated. The sum of the errors (a) and (b) was added in quadrature to compute the total error.

Measurement of mortality rates via plating—To measure and compare the effect of nitrite on mortality at pH 6.0 and 7.3, PD Nar⁺ and RH Nap⁺ were subjected to varying nitrite concentrations, and the number of viable cells over time was measured. This allowed us to infer a mortality rate as a function of pH and nitrite concentration.

First, plating media were selected to identify PD Nar⁺ and RH Nap⁺ separately. Preliminary experiments indicated that PD Nar⁺ colonies were most clearly visible and well-defined on 1/10X TSB, 1.5% agar plates (1.5 g/L tryptone, 0.5 g/L soytone, 0.5 g/L sodium chloride, 15 g/L Bacto agar), and that RH Nap⁺ colonies were most distinguishable on a defined medium with mannitol as the carbon source. Mannitol defined medium (MDM) was made using the same protocol as SDM, but 160 mM mannitol was substituted for succinate, and 1.5% w/v agar was added.

Strains were then prepared for mortality experiments by preculturing as in phenotyping experiments (i.e., two stages of aerobic growth and one stage of anaerobic growth). After the anaerobic pre-culture step, 150 μ L of culture of each strain normalized to OD₆₀₀ = 0.04 was passaged under anaerobic conditions to 1050 μ L of SDM (4 mM succinate, no nitrate) and incubated at 30 °C and 950 RPM. Two pH conditions (6 and 7.3) and four initial nitrite conditions (0, 0.438, 0.875, and 1.75 mM) were tested.

Mortality rates were measured via time course of viable cell counts in each culture. 10 μ L samples were taken from each condition at each time point. These samples were then serially tenfold diluted to achieve suspensions containing less than approximately 100 cells; the number of dilutions necessary for each strain in each condition were determined by preliminary experiments. 10 μ L of diluted samples were then pipetted onto solid medium, and streaked by tipping the plate at a roughly 45° angle until droplets approached the opposite edge of the plate. We eschewed mechanical streaking to avoid measurement bias. Plates were then covered and incubated at 30 °C. Colonies in each streak were counted manually when distinct but not overgrown (approximately 1 day for PD Nar⁺ on 1/10x TSB and 2 days for RH Nap⁺ on MDM).

We mathematically combined cell density inferences from multiple streaks to reduce measurement error. Cell density inferred from each streak at dilution level D_i can be written as

$$n_i = D_i C_i,$$

(8)

where n_i is estimated cell density and C_i is the number viable cells counted. To combine counts across dilution levels, weighted averages were used (98):

$$n_c = \frac{\sum_i \bar{n}_i w_i}{\sum_i w_i} \quad (9)$$

where $w_i = 1 / \sigma_i^2$ are the weights and $\bar{n}_i$ and σ_i are the average and standard deviation of the counts across biological replicates at dilution level i . The error on the combined counts is then given by

$$\sigma_c = \frac{1}{\sqrt{\sum_i w_i}}. \quad (10)$$

In the presence of growth, it is not possible to accurately measure mortality rate via plating. At pH 7.3, growth was observed for all supplied nitrite values $I_0 > 0$, so so these conditions were discarded from analysis. At pH 6.0, growth was observed at $I_0 = 0.5$ mM (monotonic decrease in nitrite concentration, Fig. S8), so this condition was also discarded from further analysis. For the conditions showing no growth, mortality rate was inferred using a log-linear fit. By assuming the mortality rate is proportional to the number of cells, the number of viable cells can be expressed as follows:

$$n_c = n_{c,0} \exp(-r_d t) \quad (11)$$

where r_d is the death rate, and $n_{c,0}$ is the number of viable cells at $t = 0$. Thus,

$$\ln(n_c) = \ln(n_{c,0}) - r_d t, \quad (12)$$

which is linear with respect to time. We, therefore, performed least-squares fits of $\ln(n_{c,0}) - r_d t$ to the logarithm of the measured number of viable counts and obtained $\ln(n_{c,0})$ and r_d as fit parameters.

Topsoil incubation experiments

Soil collection—To identify the taxa responsible for denitrification activity in natural soils using culture-free methods, topsoils were sampled across a range of pH values (5.0–7.1) from the Cook Agronomy Farm in September 2022 (Table S4). The Cook Agronomy Farm (CAF, 46.78°N, 117.09°W, 800 m above sea level) is a long-term agricultural research site located near Pullman, Washington, USA. CAF was established in 1998 as part of the Long-Term Agroecosystem Research (LTAR) network supported by the United States

Department of Agriculture. Before being converted to an agricultural field, the site was zonal xeric grassland or steppe. CAF operates on a continuous dryland-crop rotation system comprising winter wheat and spring crops. CAF is located in the high rainfall zone of the Pacific northwest region and the soil type is classified as Mollisol (Naff, Thatuna and Palouse Series) (99).

Ten topsoils were collected from the eastern region of the CAF at a depth of 10–20 cm. Samples were collected within a diameter of 500 m to minimize the variation of edaphic factors other than pH. The large variation of soil pH comes from the long-term use of ammoniacal fertilizers and associated N transformations, combined with field-scale hydrologic processes that occur under continuous no-tillage superimposed over a landscape that has experienced long-term soil erosion. The pH measurements were made using a glass electrode in a 1:5 (soil to water) suspension of soil in Milli-Q filtered water. The ten soils had similar edaphic properties: 6–8% gravimetric water content (g/g), soil texture of silty clay or silty clay loam with 36–41% clay, and C:N ratio constant at an approximate value of 12 with 1.1–1.8% total carbon (wt/wt) (Table S4).

Soil processing and incubation—To process soils for incubation, samples were sieved (<2 mm) to remove apparent plant roots and stones, and water content was measured (via drying at 105 °C for 24 h). To mimic autumn rainfall in the CAF area and stabilize microbial activity before beginning incubation, we rewetted the soil with sterile Milli-Q water at 40% water holding capacity for 2 weeks at room temperature. Soil slurries were then made by adding sterilized water to soil (2:1 w/w ratio of water to soil), which is close to natural state of soil saturated with water. Three replicates of each soil sample were then mixed with a concentrated sodium nitrate solution (nitrate⁺ conditions) to yield an approximate 2 mM final concentration, and another three with sterile water (controls). These slurries were then transferred to 48-deepwell plates for incubation under anaerobic conditions (950 RPM, 30 °C) for approximately 96 h. Soil extracts were then prepared using a 2 M KCl solution, followed by 0.22 µm filtering to remove soil particles that may interfere with colorimetric assays. Colorimetric assays to measure the amount of nitrate consumed were carried out using a plate reader as described in earlier experiments. Post-incubation, endpoint samples were stored at –80 °C for subsequent DNA extraction and sequencing.

Sequencing and analysis of topsoil incubations—Genomic DNA was extracted from 400 µL incubation endpoint subsamples in a combined chemical and mechanical procedure using the PowerSoil DNeasy PowerSoil HTP 96 Kit (Qiagen, Hilden, Germany). Extraction was performed following the manufacturer's protocol, and extracted DNA was stored at –20 °C. To estimate the absolute abundance of bacterial 16S rRNA amplicons, known quantities of gDNA belonging to *Escherichia coli* B and *Parabacteroides* sp. TM425 (sample obtained from Duchossois Family Institute Commensal Isolate Library, Chicago, IL, USA) were added to the slurry subsamples before the DNA extraction step. DNA Library preparation was performed using the 16S Metagenomic Sequencing Library Preparation protocol with a 2-stage PCR workflow (Illumina, San Diego, CA, United States). The V3–V4 region was amplified using forward primer 341-b-S-17 (CCTACGGGNGGCWGCAG) and reverse primer 785-a-A-21 (GACTACHVGGGTATCTAATCC) (100). We confirmed

using gel electrophoresis that the negative samples containing all reagents did not show visible bands after PCR amplification. Sequences were obtained on the Illumina MiSeq platform in a 2×300 bp paired-end run using the MiSeq Reagent Kit v3 (Illumina, San Diego, CA, United States). A standardized 10-strain gDNA mixture (MSA-1000, ATCC, Manassas, VA, USA) was sequenced as well to serve as a positive control.

Raw Illumina sequencing reads were stripped of primers, truncated at Phred quality score 2, trimmed to length 263 for forward reads and 189 for reverse reads (ensuring a 25-nucleotide overlap for most reads), and filtered to a maximum expected error of 4 based on Phred scores; this pre-processing was performed with USEARCH ver. 11.0 (101). The filtered reads were then processed with DADA2 ver. 1.26 following the developers' recommended pipeline (94). Briefly, forward and reverse reads were denoised separately, then merged and filtered for chimeras. For greater sensitivity, ASV inference was performed using the DADA2 pseudo-pooling mode, pooling samples by soil. After processing, the sequencing depth of denoised samples was 10^4 – 10^5 reads per sample. Low-abundance ASVs were dropped, retaining 4466 ASVs for further analysis. (The samples used in this work were collected as part of a larger study; the exact abundance-filtering criterion was to retain ASVs with cumulative abundance of at least 1000 counts across the 902 samples of this larger dataset.) Taxonomy was assigned by DADA2 using the SILVA database ver. 138.1, typically at genus level, but with species-level attribution recorded in cases of a 100% sequence match.

As an internal control, we verified that the ASVs corresponding to the two spiked-in genera *Escherichia-Shigella* and *Parabacteroides* were highly correlated with each other as expected (Pearson correlation $\rho = 0.94$). These ASVs were removed from the table and combined into a single reference vector of “spike-in counts”. The spike-in counts constituted $5.5 \pm 2.5\%$ of total reads in each sample.

For subsequent analysis, the raw ASV counts were augmented by a pseudocount of 0.5 and divided by the per-sample spike-in counts, yielding values that can be interpreted as the absolute biomass of each taxon (up to a factor corresponding to the copy number of the 16S operon), measured in units where 1 means as many 16S fragments as the number of DNA molecules in the spike-in.

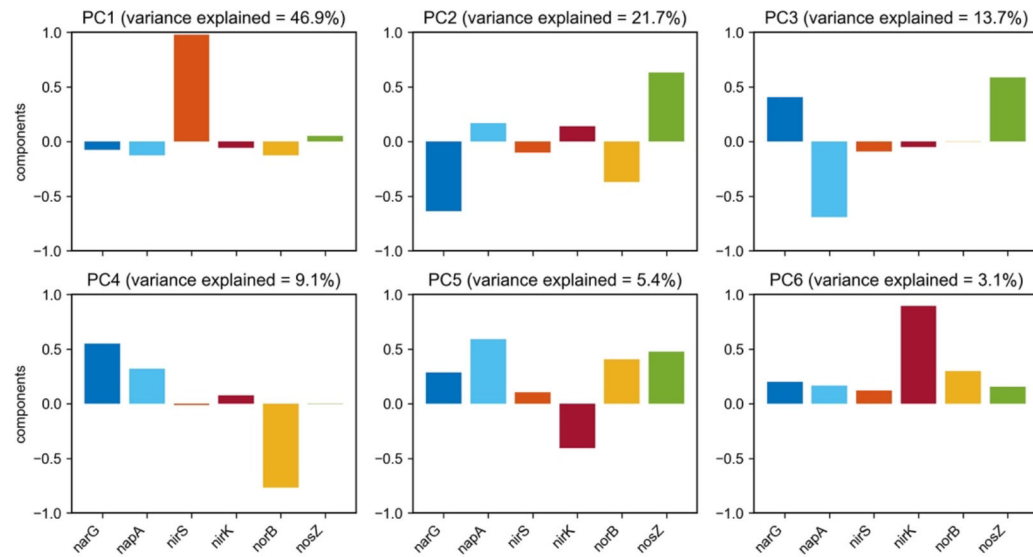
Identifying ASVs enriched on nitrate in topsoil incubations—To identify the ASVs enriched in nitrate treatments versus the no-nitrate controls, it was necessary to determine what change in recorded abundance constitutes a significant change, relative to what might be expected for purely stochastic reasons. The relevant null model would combine sampling and sequencing noise with the stochasticity of ecological dynamics over a 4-day incubation, and cannot be derived from first principles. However, since all measurements were performed in triplicate with independent incubations, the relevant null model can be determined empirically. The deviations of replicate-replicate comparisons from 1:1 line were well-described by an effective model combining two independent contributions, a Gaussian noise of fractional magnitude c_{frac} and a constant Gaussian noise of magnitude c_0 reads, such that repeated measurements (over biological replicates) of an ASV with mean abundance n counts are approximately Gaussian-distributed with a

standard deviation of $\sigma(c_0, c_{\text{frac}}) = \sqrt{(c_{\text{frac}}n)^2 + c_0^2}$ counts (Fig. S9). In this expression, c_{frac} was estimated from moderate-abundance ASVs (> 50 counts) for which the other noise term is negligible; and c_0 was then determined as the value for which 67% of replicate-replicate comparisons are within $\pm\sigma(c_0, c_{\text{frac}})$ of each other, as expected for 1-sigma deviations. This noise model was inferred separately for each soil, as the corresponding samples were processed independently in different sequencing runs; the parameters across 10 soils were $c_{\text{frac}} = 0.22 \pm 0.03$ and $c_0 = 11 \pm 3$ counts. The model was used to compute the z-scores for the enrichments of absolute ASV abundances in nitrate treatments against no-nitrate controls (three independent z-scores, from triplicate treatments). Significantly enriched ASVs were identified in each sample as those with z-scores greater than $z = \Phi^{-1}(1 - \alpha / 2 / n_{\text{ASV}})$, where $\Phi^{-1}(x)$ is the inverse CDF of the standard normal distribution, $\alpha = 0.01$, and n_{ASV} as the number of nonzero ASVs in a given sample. This critical z-score ($z = 4.51 \pm 0.04$) corresponds to a two-tailed Bonferroni-corrected hypothesis test at significance level α under the null hypothesis that counts in the nitrate+ and control conditions are drawn from the same distribution. These analyses were performed using custom MATLAB scripts (Mathworks, Inc), which are available on the OSF data repository for the present manuscript; for additional technical details, the reader is referred to the detailed comments in these scripts.

In order to determine which significantly enriched ASVs are likely to represent denitrifiers, PICRUSt2 ver. 2.5.2 (102) was used to assign putative genotypes. PICRUSt2 matches input 16S rRNA sequences to genotypes using a curated reference genome database. A list of inferred KEGG orthologs for each ASV was computed using the default parameters of the PICRUSt2 pipeline. An ASV was classified as a denitrifier if KEGG orthologs for any denitrification reductase (*narG*/K00370, *napA*/K02567, *nirS*/K15864, *nirK*/K00368, *norB*/K04561, *nosZ*/K00376) were inferred *and* if the DNRA nitrite reductase *nrfA*/K03385 was not inferred. ASVs were classified as performing DNRA if *nrfA*/K03385 was inferred, and as neither performing denitrification nor DNRA otherwise.

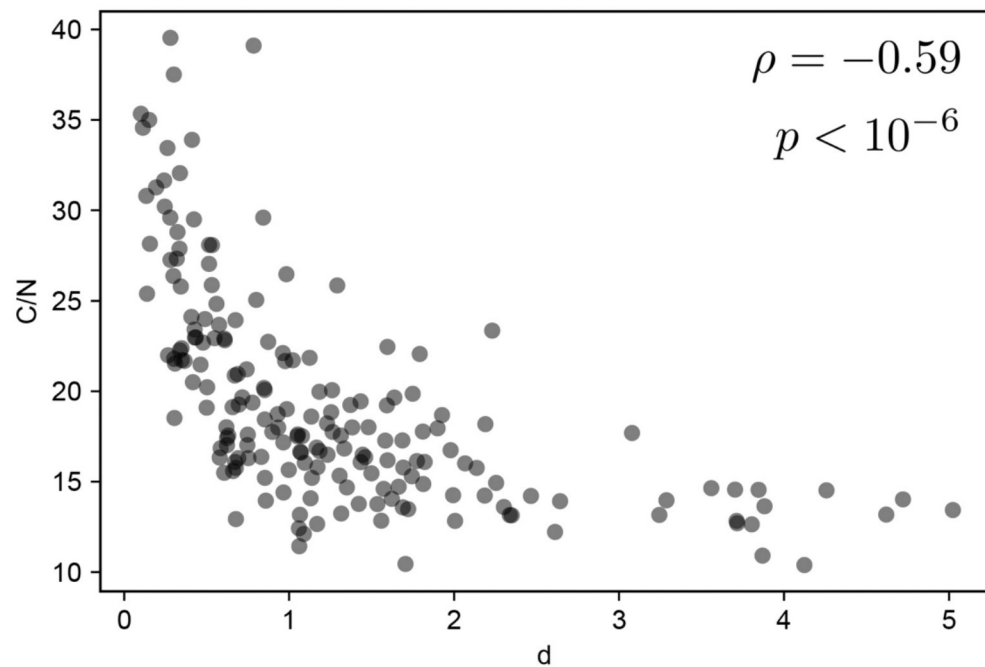
Statistics & Reproducibility—No data were excluded from the analyses.

Extended Data



Extended Data Figure 1:

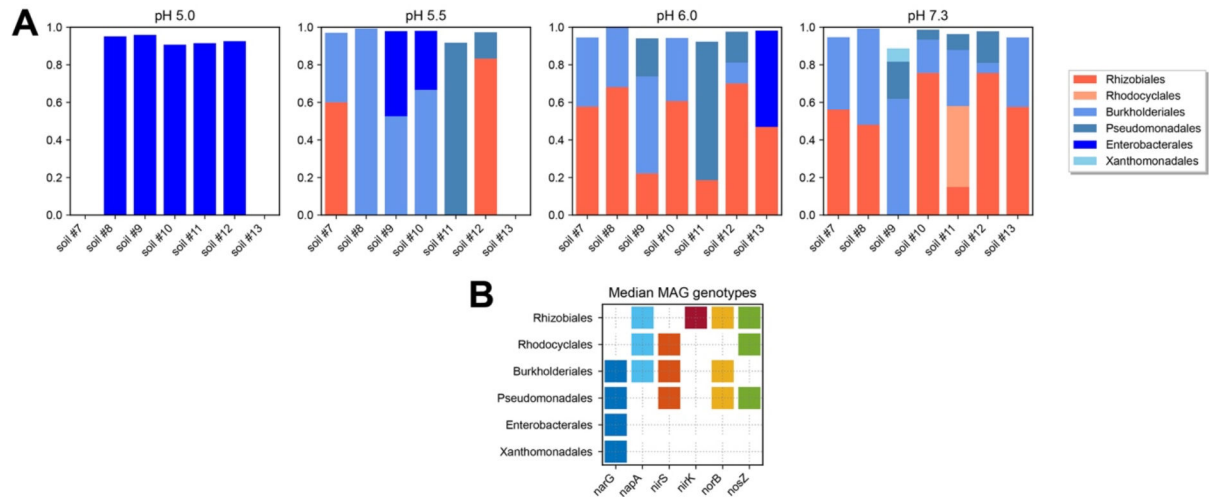
Bars show the loadings of each gene in the six principal components (PCs) resulting from the uiSVD decomposition of denitrification reductase (*narG*, *napA*, *nirS*, *nirK*, *norB*, *nosZ*) relative abundances from the global topsoil microbiome.



Extended Data Figure 2:

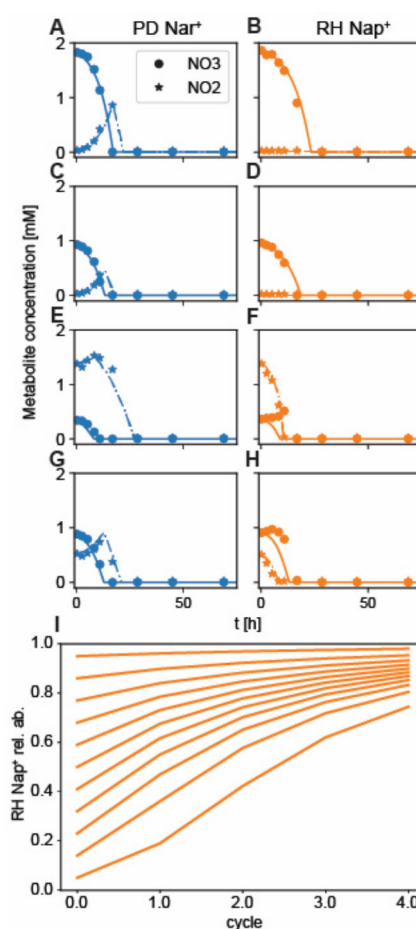
Unit-invariant singular value decomposition (uiSVD) was used to decompose denitrification reductase (*narG*, *napA*, *nirS*, *nirK*, *norB*, *nosZ*) relative abundances from the global topsoil microbiome into contributions due to pathway magnitude and composition (Fig. 1A-D).

Pathway magnitude (*d*) most strongly correlated with C/N ratio ($\rho = -0.59$, $p < 10^{-6}$ via one-tailed randomization test; Fig. 1F).



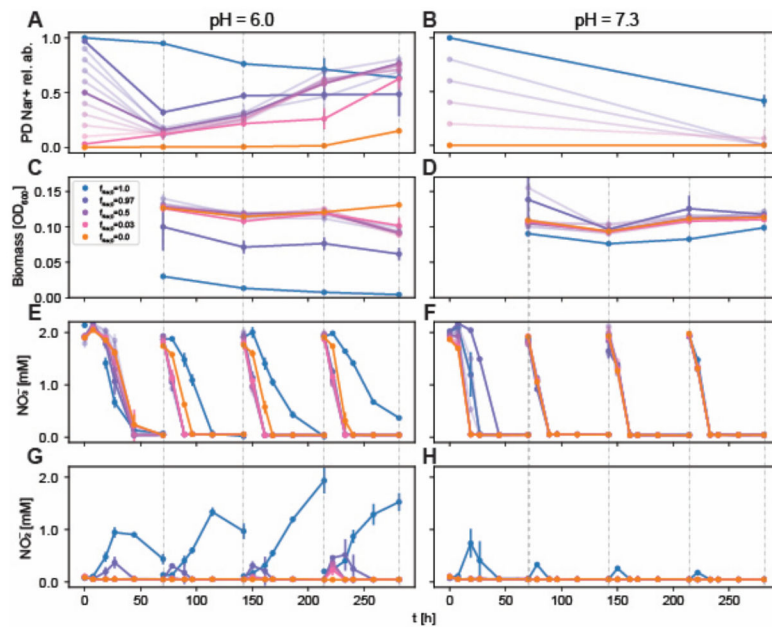
Extended Data Figure 3:

Additional enrichments were performed at pH 5.0, 5.5, 6.0, and 7.3 and the endpoint cultures were shotgun sequenced to infer taxonomic composition and genotypes. **(A)** Endpoint community compositions of the enrichments inferred via 16S miTAGs are shown. Taxa with Nar⁺ genotypes are indicated in shades of blue, while taxa that possess Nap and not Nar are indicated in shades of red. Compositions are shown at the level of taxonomic order, and taxa present at a level of less than 1% are omitted. **(B)** Median denitrification reductase genotypes inferred via annotation of metagenome assembled genomes are shown.



Extended Data Figure 4:

(A)-(H) Consumer resource model (Supplementary Information Eq. S1) fit to monoculture metabolite data at pH 7.3. Dots indicate nitrate concentrations, stars indicate nitrite concentrations, solid lines show fits to nitrate dynamics, and dash-dot lines show fits to nitrite dynamics. All concentrations are averaged over biological replicates ($n = 3$). Panels (A, C, E, G) are fits to PD Nar^+ data, while panels (B, D, F, H) are fits to RH Nap^+ data. To infer nitrate and nitrite reduction rates independently fits were performed for a number of different initial conditions ($[\text{NO}_3^-]$, $[\text{NO}_2^-]$). Panels (A) and (B) correspond to (1.75, 0), (C) and (D) to (0.875, 0), (E) and (F) to (0.4375, 1.3125), (G) and (H) to (0.875, 0.4375). All concentrations are reported in units of mM. (I) Co-culture relative abundance prediction based on monoculture phenotypes. RH Nap^+ is predicted to approach a relative abundance of 1 for all initial conditions at pH 7.3. This is consistent with what was observed in enrichment experiments (Fig. 5F).



Extended Data Figure 5:

Details of the PD Nar^+ and RH Nar^+ co-culture experiment shown in Fig. 5. **(A)** PD Nar^+ relative abundance dynamics at pH 6.0 are shown. $f_{0, \text{Nar}^+} = 0, 0.03, 0.5, 0.97$ and 1 are highlighted. Due to small levels of cross-contamination between pure and mixed cultures, f_{Nar^+} increases from 0 and decreases from 1. Although this was unintentional, it indicates that each of these strains is inviable by the other in this condition, providing more evidence that they coexist. **(B)** PD Nar^+ relative abundance dynamics at pH 7.3 are shown. In panels A and B, data points are means of biological replicates ($n = 4$ for PD Nar^+ relative abundance ≥ 0.5 at pH 6 and relative abundance ≤ 0.03 and $= 1$ in both pH conditions; $n = 3$ for all other conditions) of inferred relative abundances (Methods) and errorbars are calculated as described in the Fig. 5 caption and Methods. **(C)** Endpoint biomass dynamics, measured via absorbance at 600 nm, are shown for each cycle at pH 6.0. The $f_{0, \text{Nar}^+} = 0$ condition produces much less biomass than the other conditions, as expected. **(D)** Endpoint biomass dynamics are shown for each cycle at pH 7.3. **(E, G)** NO_3^- and NO_2^- dynamics are measured using a Griess assay (117) and shown at pH 6.0. Aside from the $f_{0, \text{Nar}^+} = 0$ condition, for which biomass is very low (panel C), increasing f_{Nar^+} (panel A) corresponds to increasing nitrite accumulation. **(F, H)** Metabolite dynamics are shown at pH 7.3. Decreasing f_{Nar^+} (B) corresponds to decreasing nitrite accumulation. Points and error bars in panels C-H show means and standard deviations over biological replicates ($n = 4$ for PD Nar^+ relative abundance ≥ 0.5 at pH 6 and relative abundance ≤ 0.03 and $= 1$ in both pH conditions; $n = 3$ for all other conditions) in each condition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We acknowledge Luis Miguel de Jesús Astacio and Mahmoud Yousef for collecting soils, Rita Almeida Oliveira for help with colony counting technique, Mohammad Bahram for previously published global topsoil microbiome data, and Kabir Husain for useful discussions. Soil sampling done in the Cook Agronomy Farm was, in part, a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture.

This work was supported by the National Science Foundation Division of Emerging Frontiers EF 2025293 (S.K.) and EF 2025521 (M.M.), as well as, the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1746045 (M.C-W). S.K. acknowledges the Center for the Physics of Evolving Systems at the University of Chicago, National Institute of General Medical Sciences R01GM151538, and support from the National Science Foundation through the Center for Living Systems (grant no. 2317138). K.G. acknowledges a James S. McDonnell Foundation Postdoctoral Fellowship Award 220020499. M.T. acknowledges support from NSF grant PHY-2310746. M.C-W. acknowledges a Fannie and John Hertz Fellowship Award. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References and Notes

1. Falkowski PG, Fenchel T & Delong EF The microbial engines that drive Earth's biogeochemical cycles. *Science* 320, 1034–1039 (2008). URL 10.1126/science.1153213. [PubMed: 18497287]
2. Song W et al. Functional traits resolve mechanisms governing the assembly and distribution of nitrogen-cycling microbial communities in the global ocean. *mBio* 13, e03832–21 (2022). [PubMed: 35285696]
3. Wang C.-y. et al. Soil pH is the primary factor driving the distribution and function of microorganisms in farmland soils in northeastern China. *Annals of Microbiology* 69, 1461–1473 (2019).
4. Bahram M et al. Structure and function of the global topsoil microbiome. *Nature* 560, 233–237 (2018). URL 10.1038/s41586-018-0386-6. [PubMed: 30069051]
5. Malik AA, Thomson BC, Whiteley AS, Bailey M & Griffiths RI Bacterial physiological adaptations to contrasting edaphic conditions identified using landscape scale metagenomics. *mBio* 8, e00799–17 (2017). [PubMed: 28679747]
6. Sofi MH et al. pH of drinking water influences the composition of gut microbiome and type 1 diabetes incidence. *Diabetes* 63, 632–644 (2014). [PubMed: 24194504]
7. Zinöcker MK & Lindseth IA The western diet–microbiome–host interaction and its role in metabolic disease. *Nutrients* 10, 365 (2018). [PubMed: 29562591]
8. Singh RK et al. Influence of diet on the gut microbiome and implications for human health. *Journal of Translational Medicine* 15, 1–17 (2017). [PubMed: 28049494]
9. Baas-Becking LGM Geobiologie of Inleiding Tot de Milieukunde (Van Stockkum Zoon, The Hague, 1934).
10. Dimitriu PA & Grayston SJ Relationship between soil properties and patterns of bacterial β -diversity across reclaimed and natural boreal forest soils. *Microbial Ecology* 59, 563–573 (2010). [PubMed: 19830478]
11. Rousk J et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal* 4, 1340–1351 (2010). [PubMed: 20445636]
12. Jones RT et al. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *The ISME Journal* 3, 442–453 (2009). [PubMed: 19129864]
13. Faust K & Raes J Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10, 538–550 (2012). [PubMed: 22796884]
14. Pacheco AR & Segrè D A multidimensional perspective on microbial interactions. *FEMS Microbiology Letters* 366, fnz125 (2019). [PubMed: 31187139]
15. Gopalakrishnappa C, Li Z & Kuehn S Environmental modulators of algae-bacteria interactions at scale. *bioRxiv* (2023).
16. Pearl Mizrahi S, Goyal A & Gore J Community interactions drive the evolution of antibiotic tolerance in bacteria. *Proceedings of the National Academy of Sciences* 120, e2209043119 (2023).

URL <https://www.pnas.org/doi/10.1073/pnas.2209043119>. Publisher: Proceedings of the National Academy of Sciences.

17. Hsu RH et al. Microbial interaction network inference in microfluidic droplets. *Cell Systems* 9, 229–242 (2019). [PubMed: 31494089]
18. Sherr EB & Sherr BF Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* 81, 293–308 (2002). [PubMed: 12448728]
19. de la Cruz Barron M et al. Shifts from cooperative to individual-based predation defense determine microbial predator-prey dynamics. *The ISME Journal* 1–11 (2023). [PubMed: 36127432]
20. Granato ET, Meiller-Legrand TA & Foster KR The evolution and ecology of bacterial warfare. *Current Biology* 29, R521–R537 (2019). [PubMed: 31163166]
21. Venturelli OS et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Molecular Systems Biology* 14, e8157 (2018). [PubMed: 29930200]
22. Gamble TN, Betlach MR & Tiedje JM Numerically Dominant Denitrifying Bacteria from World Soils 1. *Applied and Environmental Microbiology* 33, 926–939 (1977). URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC170793/>. [PubMed: 869539]
23. Seitzinger S et al. Denitrification across landscapes and waterscapes: A synthesis. *Ecological Applications* 16, 2064–2090 (2006). [PubMed: 17205890]
24. Lu H, Chandran K & Stensel D Microbial ecology of denitrification in biological wastewater treatment. *Water Research* 64, 237–254 (2014). URL 10.1016/j.watres.2014.06.042. [PubMed: 25078442]
25. Irazábal T, Belcheva A, Girardin SE, Martin A & Philpott DJ The multifaceted role of the intestinal microbiota in colon cancer. *Molecular Cell* 54, 309–320 (2014). [PubMed: 24766895]
26. Zumft WG Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews* 61, 533–616 (1997). URL 10.1128/mmbr.61.4.533-616.1997. [PubMed: 9409151]
27. Rodionov DA, Dubchak IL, Arkin AP, Alm EJ & Gelfand MS Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS computational biology* 1 (2005).
28. Shapleigh JP The denitrifying prokaryotes. In *The Prokaryotes*, 769–792 (2006).
29. Heylen K et al. Cultivation of denitrifying bacteria: Optimization of isolation conditions and diversity study. *Applied and Environmental Microbiology* 72, 2637–2643 (2006). [PubMed: 16597968]
30. Lycus P et al. Phenotypic and genotypic richness of denitrifiers revealed by a novel isolation strategy. *ISME Journal* 11, 2219–2232 (2017). URL 10.1038/ismej.2017.82. [PubMed: 28696424]
31. Gowda K, Ping D, Mani M & Kuehn S Genomic structure predicts metabolite dynamics in microbial communities. *Cell* 185, 530–546 (2022). [PubMed: 35085485]
32. Uhlmann J Unit Consistency, Generalized Inverses, and Effective System Design Methods (2016). URL <http://arxiv.org/abs/1604.08476>. 1604.08476.
33. Lauber CL, Hamady M, Knight R & Fierer N Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* 75, 5111–5120 (2009). [PubMed: 19502440]
34. Šimek M & Cooper JE The influence of soil pH on denitrification: Progress towards the understanding of this interaction over the last 50 years. *European Journal of Soil Science* 53, 345–354 (2002).
35. Lilja EE & Johnson DR Segregating metabolic processes into different microbial cells accelerates the consumption of inhibitory substrates. *ISME Journal* 10, 1568–1578 (2016). URL 10.1038/ismej.2015.243. [PubMed: 26771930]
36. Tiedje JM, Sextstone AJ, Myrold DD & Robinson JA Denitrification: ecological niches, competition and survival. *Antonie van Leeuwenhoek* 48, 569–583 (1983). URL 10.1007/BF00399542.
37. Magdoff F & Bartlett R Soil ph buffering revisited. *Soil Science Society of America Journal* 49, 145–148 (1985).

38. Nelson PN & Su N Soil ph buffering capacity: a descriptive function and its application to some acidic tropical soils. *Soil Research* 48, 201–207 (2010).
39. Bache B The role of calcium in buffering soils. *Plant, Cell & Environment* 7, 391–395 (1984).
40. Kreitler CW & Jones DC Natural soil nitrate: The cause of the nitrate contamination of ground water in runnels county, texas a. *Groundwater* 13, 53–62 (1975).
41. Dowdell RJ & Cannell RQ Effect of Ploughing and Direct Drilling on Soil Nitrate Content. *Journal of Soil Science* 26 (1975). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2389.1975.tb01929.x>.
42. Uritskiy GV, Diruggiero J & Taylor J MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, 1–13 (2018). [PubMed: 29291746]
43. Bollag J-M & Henninger NM Effects of nitrite toxicity on soil bacteria under aerobic and anaerobic conditions. *Soil Biology and Biochemistry* 10, 377–381 (1978).
44. Tarr H The action of nitrites on bacteria. *Journal of the Fisheries Board of Canada* 5, 265–275 (1941).
45. Bancroft K, Grant IF & Alexander M Toxicity of no₂: effect of nitrite on microbial activity in an acid soil. *Applied and Environmental Microbiology* 38, 940–944 (1979). [PubMed: 16345466]
46. Durvasula K et al. Effect of periplasmic nitrate reductase on diauxic lag of paracoccus pantotrophus. *Biotechnology progress* 25, 973–979 (2009). [PubMed: 19399903]
47. Richardson DJ & Ferguson SJ The influence of carbon substrate on the activity of the periplasmic nitrate reductase in aerobically grown thiosphaera pantotropha. *Archives of microbiology* 157, 535–537 (1992).
48. Wang H, Tseng CP & Gunsalus RP The napF and narG nitrate reductase operons in Escherichia coli are differentially expressed in response to submicromolar concentrations of nitrate but not nitrite. *Journal of Bacteriology* 181, 5303–5308 (1999). [PubMed: 10464201]
49. Potter LC, Millington P, Griffiths L, Thomas GH & Cole JA Competition between Escherichia coli strains expressing either a periplasmic or a membrane-bound nitrate reductase: Does Nap confer a selective advantage during nitrate-limited growth? *Biochemical Journal* 344, 77–84 (1999). [PubMed: 10548536]
50. Kraft NJ et al. Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* 29, 592–599 (2015).
51. Abraham ZH, Lowe DJ & Smith BE Purification and characterization of the dissimilatory nitrite reductase from Alcaligenes xylosoxidans subsp. Xylosoxidans (N.C.I.M.B. 11015): Evidence for the presence of both type 1 and type 2 copper centres. *Biochemical Journal* 295, 587–593 (1993). [PubMed: 8240262]
52. Masuko M, Iwasaki H, Sakurai T, Suzuki S & Nakahara A Characterization of nitrite reductase from a denitrifier, alcaligenes sp. ncib 11015. a novel copper protein. *The Journal of Biochemistry* 96, 447–454 (1984). [PubMed: 6501252]
53. Iwasaki H & Matsubara T A nitrite reductase from Achromobacter cycloclastes. *Journal of Biochemistry* 71, 645–652 (1972). [PubMed: 5042456]
54. Liu MY, Liu MC, Payne WJ & Legall J Properties and electron transfer specificity of copper proteins from the denitrifier ‘Achromobacter cycloclastes’. *Journal of Bacteriology* 166, 604–608 (1986). [PubMed: 3700338]
55. Kakutani T, Watanabe H, Arima K & Beppu T Purification and properties of a copper-containing nitrite reductase from a denitrifying bacterium, Alcaligenes faecalis strain S-6. *Journal of Biochemistry* 89, 453–461 (1981). [PubMed: 7240122]
56. Kukimoto M et al. X-ray Structure and Site-Directed Mutagenesis of a Nitrite Reductase from Alcaligenes faecalis S-6: Roles of Two Copper Atoms in Nitrite Reduction. *Biochemistry* 33, 5246–5252 (1994). [PubMed: 8172899]
57. Michalski WP Molecular characterization of a copper-containing nitrite reductase from rhodopseudomonas sphaeroides forma sp. denitrificans. *Biochim. Biophys. Acta* 828, 130–137 (1985).
58. Sawada E, Satoh T & Kitamura H Purification and properties of a dissimilatory nitrite reductase of a denitrifying phototrophic bacterium. *Plant and Cell Physiology* 19, 1339–1351 (1978).

59. Denariáz G, Payne WJ & LeGall J The denitrifying nitrite reductase of *Bacillus halo-denitrificans*. *BBA - Bioenergetics* 1056, 225–232 (1991).
60. Glockner AB, Jüngst A & Zumft WG Copper-containing nitrite reductase from *Pseudomonas aureofaciens* is functional in a mutationally cytochrome cd1-free background (NirS-) of *Pseudomonas stutzeri*. *Archives of Microbiology* 160, 18–26 (1993). [PubMed: 8352648]
61. Timkovich R, Dhesi R, Martinkus KJ, Robinson MK & Rea TM Isolation of *Paracoccus denitrificans* cytochrome cd1: Comparative kinetics with other nitrite reductases. *Archives of Biochemistry and Biophysics* 215, 47–58 (1982). [PubMed: 6284044]
62. Gordon EH et al. Structure and kinetic properties of *Paracoccus pantotrophus* cytochrome cd1 nitrite reductase with the d1 heme active site ligand tyrosine 25 replaced by serine. *Journal of Biological Chemistry* 278, 11773–11781 (2003). [PubMed: 12556530]
63. Besson S, Carneiro C, Moura JJ, Moura I & Fauque G A cytochrome cd1-type nitrite reductase isolated from the marine denitrifier *pseudomonas nautica* 617: Purification and characterization. *Anaerobe* 1, 219–226 (1995). [PubMed: 16887530]
64. Sawhney V & Nicholas DJ Sulphide-linked nitrite reductase from *Thiobacillus denitrificans* with cytochrome oxidase activity: purification and properties. *Journal of General Microbiology* 106, 119–128 (1978).
65. Gowda K, Ping D, Mani M & Kuehn S Genomic structure predicts metabolite dynamics in microbial communities. *Cell* 185, 530–546 (2022). URL 10.1016/j.cell.2021.12.036. [PubMed: 35085485]
66. Stewart V, Lu Y & Darwin AJ Periplasmic nitrate reductase (NapABC enzyme) supports anaerobic respiration by *Escherichia coli* K-12. *Journal of Bacteriology* 184, 1314–1323 (2002). [PubMed: 11844760]
67. Van Alst NE, Sherrill LA, Iglewski BH & Haidaris CG Compensatory periplasmic nitrate reductase activity supports anaerobic growth of *Pseudomonas aeruginosa* PAO1 in the absence of membrane nitrate reductase. *Canadian Journal of Microbiology* 55, 1133–1144 (2009). [PubMed: 19935885]
68. Ikeda E et al. Physiological roles of two dissimilatory nitrate reductases in the deep-sea denitrifier *Pseudomonas* sp. strain MT-1. *Bioscience, Biotechnology and Biochemistry* 73, 896–900 (2009). [PubMed: 19352028]
69. Bell LC, Richardson DJ & Ferguson SJ Periplasmic and membrane-bound respiratory nitrate reductases in *Thiosphaera pantotropha*. The periplasmic enzyme catalyzes the first step in aerobic denitrification. *FEBS Letters* 265, 85–87 (1990). [PubMed: 2365057]
70. Warnecke-Eberz U & Friedrich B Three nitrate reductase activities in *Alcaligenes eutrophus*. *Archives of Microbiology* 159, 405–409 (1993).
71. Rigoulet M, Mourier A, Galinier A, Casteilla L & Devin A Electron competition process in respiratory chain: regulatory mechanisms and physiological functions. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1797, 671–677 (2010). [PubMed: 20117078]
72. Menzel P, Ng KL & Krogh A Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7 (2016).
73. Logares R et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology* 16, 2659–2671 (2014). [PubMed: 24102695]
74. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
75. Bushnell B, Rood J & Singer E Bbmerge—accurate paired shotgun read merging via overlap. *PloS One* 12 (2017).
76. Cole JR et al. The ribosomal database project (rdp-ii): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Research* 33, D294–D296 (2005). [PubMed: 15608200]
77. Nurk S, Meleshko D, Korobeynikov A & Pevzner PA MetaSPAdes: A new versatile metagenomic assembler. *Genome Research* 27, 824–834 (2017). [PubMed: 28298430]
78. Alneberg J et al. Binning metagenomic contigs by coverage and composition. *Nature Methods* 11, 1144–1146 (2014). [PubMed: 25218180]
79. Wu Y-W, Simmons BA & Singer SW Maxbin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607 (2016). [PubMed: 26515820]

80. Kang DD, Froula J, Egan R & Wang Z MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015, 1–15 (2015).
81. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25, 1043–1055 (2015). [PubMed: 25977477]
82. Meyer F et al. The metagenomics rast server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9, 1–8 (2008). [PubMed: 18173834]
83. Quast C et al. The silva ribosomal rna gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, D590–D596 (2012). [PubMed: 23193283]
84. Glöckner FO et al. 25 years of serving the community with ribosomal rna gene reference databases and tools. *Journal of Biotechnology* 261, 169–176 (2017). [PubMed: 28648396]
85. Yilmaz P et al. The silva and “all-species living tree project (ltp)” taxonomic frameworks. *Nucleic Acids Research* 42, D643–D648 (2014). [PubMed: 24293649]
86. Waki T, Murayama K, Kawato Y & Ichikawa K Transient characteristics of paracoccus denitrificans with changes between aerobic and anaerobic conditions. *Journal of Fermentation Technology* 58 (1980).
87. Liu P-H, Zhan G, Svoronos SA & Koopman B Diauxic lag from changing electron acceptors in activated sludge treatment. *Water Research* 32, 3452–3460 (1998).
88. Liu P-H, Svoronos SA & Koopman B Experimental and modeling study of diauxic lag of pseudomonas denitrificans switching from oxic to anoxic conditions. *Biotechnology and Bioengineering* 60, 649–655 (1998). [PubMed: 10099474]
89. Gouw M, Bozic R, Koopman B & Svoronos SA Effect of nitrate exposure history on the oxygen/nitrate diauxic growth of pseudomonas denitrificans. *Water Research* 35, 2794–2798 (2001). [PubMed: 11456182]
90. Lisbon K, McKean M, Shekar S, Svoronos SA & Koopman B Effect of do on oxic/anoxic diauxic lag of pseudomonas denitrificans. *Journal of Environmental Engineering* 128, 391–394 (2002).
91. Casasús AI, Hamilton RK, Svoronos SA & Koopman B A simple model for diauxic growth of denitrifying bacteria. *Water Research* 39, 1914–1920 (2005). [PubMed: 15899290]
92. 16s metagenomic sequencing library preparation. Tech. Rep., Illumina, Inc., San Diego, CA, USA (2013).
93. Zhang J, Kobert K, Flouri T & Stamatakis A Pear: a fast and accurate illumina paired-end read merger. *Bioinformatics* 30, 614–620 (2014). [PubMed: 24142950]
94. Callahan BJ et al. Dada2: High-resolution sample inference from illumina amplicon data. *Nature Methods* 13, 581–583 (2016). [PubMed: 27214047]
95. Bokulich NA et al. Optimizing taxonomic classification of marker-gene amplicon sequences with qiime 2’s q2-feature-classifier plugin. *Microbiome* 6, 1–17 (2018). [PubMed: 29291746]
96. McLaren MR, Willis AD & Callahan BJ Consistent and correctable bias in metagenomic sequencing experiments. *Elife* 8, e46923 (2019). [PubMed: 31502536]
97. Ku HH et al. Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards* 70 (1966).
98. Taylor JR *An Introduction to Error Analysis (Second Edition)* (University Science Books, 1997).
99. Staff SS *Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. Agriculture handbook* 436 (1999).
100. Klindworth A et al. Evaluation of general 16s ribosomal rna gene pcr primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research* 41, e1–e1 (2013).
101. Edgar RC Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461 (2010). [PubMed: 20709691]
102. Douglas GM et al. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology* 38, 685–688 (2020). URL <https://www.nature.com/articles/s41587-020-0548-6>.

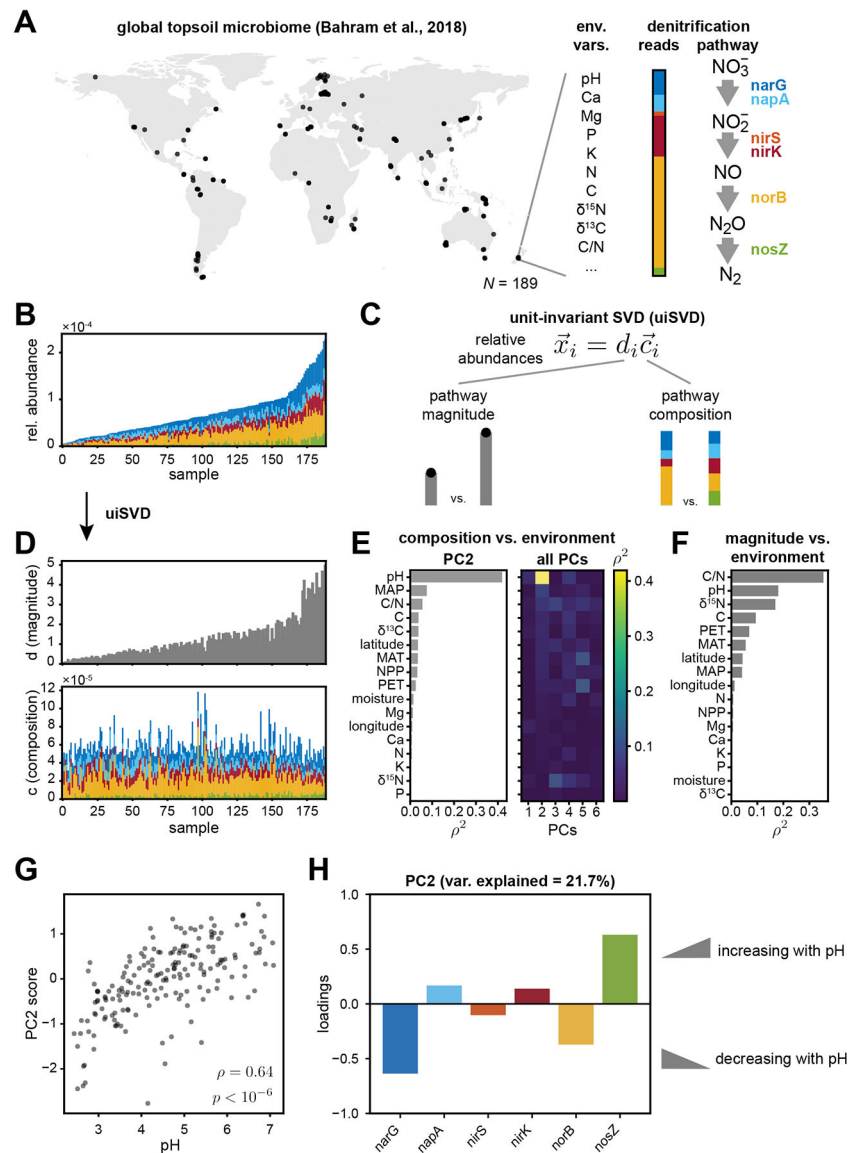


Figure 1: pH is associated with covariation in denitrification pathway composition in the global topsoil microbiome.

(A) Topsoils sampled at $n = 189$ globally-distributed sites were chemically characterized (pH, Ca, Mg, ...), sequenced via shotgun metagenomics, and functionally annotated by Bahram et al. (4). (B) The relative abundance of denitrification reductases in each soil sample (relative to total gene content) are plotted in order of increasing total relative abundance. Reductase color legend indicated in panel A. (C-D) Unit-invariant singular value decomposition (uiSVD) (32) was used to decompose the data in panel B into contributions due to pathway magnitude (d_i) and pathway composition ($\vec{c}_i$). The results of this decomposition are plotted in panel D. (E-F) Principal component (PC) scores for pathway composition and pathway magnitudes obtained via uiSVD (Methods) are compared with 17 environmental variables, and squared Pearson correlation coefficients are shown. Scores of PC2 are most correlated with pH ($\rho = 0.64$, $p < 10^{-6}$ via one-tailed randomization

test; panel G), while pathway magnitude is most correlated with C/N ratio ($\rho = -0.59$, $p < 10^{-6}$ via one-tailed randomization test; Extended Data Fig. 2). **(H)** Loadings of PC2 are shown, where positive values indicate reductase content that increases with pH, and vice versa. See also Extended Data Figs. 1, 2.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

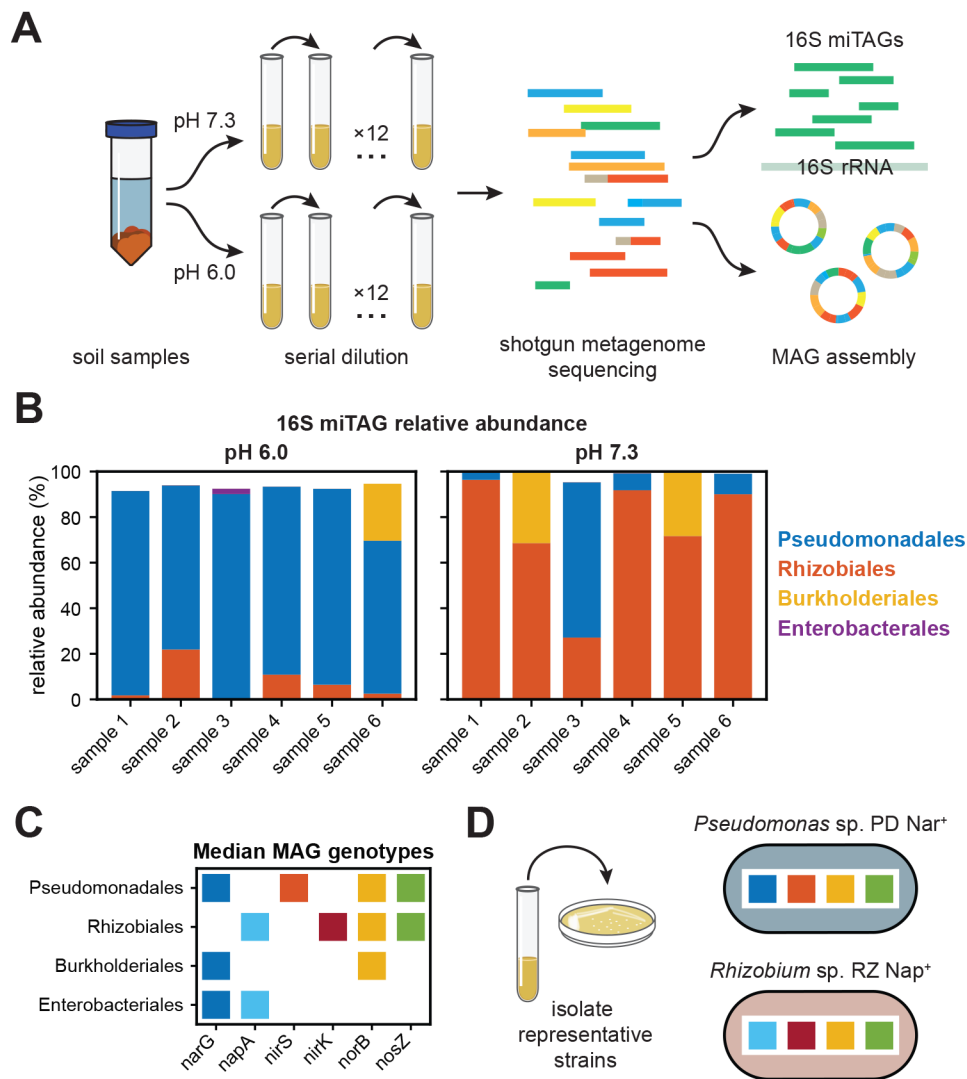


Figure 2: Enrichment cultures reproduce patterns in denitrification gene content in the topsoil microbiome.

(A) Denitrifying communities were enriched from soil samples in two pH conditions (6.0 and 7.3), and then sequenced via shotgun metagenomics to measure taxonomic composition and genotypes. Six soil samples were mechanically homogenized and used to inoculate a serial dilution experiment in denitrifying (anaerobic) conditions. After 72 h growth cycles, cultures were repeatedly passaged (12×) into a defined medium containing 1.75 mM nitrate via a 1/8 dilution factor (Methods). (B) Endpoint community compositions are shown at the level of taxonomic order for enrichments in pH 6.0 and 7.3. Compositions were determined by extracting 16S rRNA fragments (miTAGs) from shotgun metagenome data (73), and then taxonomically annotating miTAGs via RDP (76). Taxa present at a relative abundance less than 1% are omitted. (C) Median denitrification reductase genotypes are shown for metagenome-assembled genomes (MAGs) corresponding to the four most abundant taxa present in the composition data in panel B. MAGs were functionally annotated using RAST (82), and the median genotype was computed over MAGs obtained in different samples. (D) Strains were isolated from cryopreserved samples from the enrichment endpoint (Methods).

The strain *Pseudomonas* sp. PD Nar⁺ represents the dominant taxa present across samples at pH 6.0, and *Rhizobiales* sp. RH Nap⁺ represents the dominant taxa across samples at pH 7.3. See also Fig. S6, Extended Data Fig. 3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

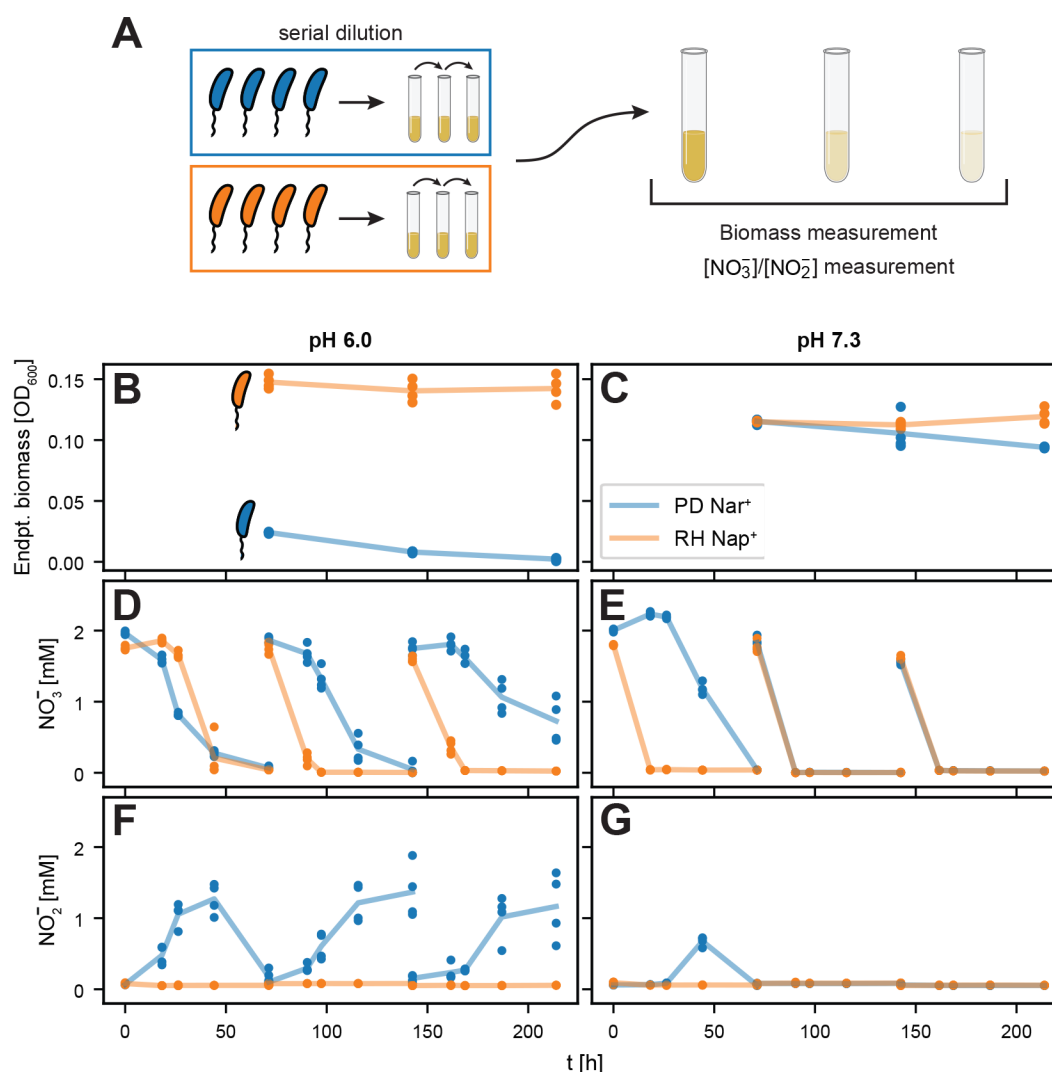


Figure 3: Individual traits do not explain the outcome of acidic enrichments.

(A) Schematic of the experimental design. PD Nar⁺ (blue) and RH Nap⁺ (orange) were passaged repeatedly in monoculture under denitrifying (anaerobic) conditions. After each 72 h growth cycle, cultures were passaged (3×) into a defined medium containing 1.75 mM nitrate using a 1/8 dilution factor. Biomass is measured via 600 nm absorbance at the end of each cycle, and nitrate and nitrite concentrations are measured throughout each cycle using a Griess assay (Methods). (B and C) Endpoint biomass is shown for each strain at both pH 6.0 (panel B) and 7.3 (panel C). Lines connect the average across replicates. RH Nap⁺ (orange) produces more biomass at both pH levels, while PD Nar⁺ (blue) appears to decay in abundance at pH 6.0 (despite PD Nar⁺ community enrichment in this condition, Fig. 2B). (D-G) Nitrate and nitrite concentration dynamics are shown at six timepoints throughout each cycle in each condition. PD Nar⁺ nitrate and nitrite reduction rates slow at pH 6.0 as the growth-dilution cycles progress, consistent with its reduction in biomass (Panels D & F, blue). PD Nar⁺ (blue) accumulates nitrite at each pH condition, whereas RH Nap⁺ (orange)

does not (Panels F & G). Biological replicates ($n = 4$) are shown for each strain in each experimental condition, with lines connecting the averages of these replicates.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

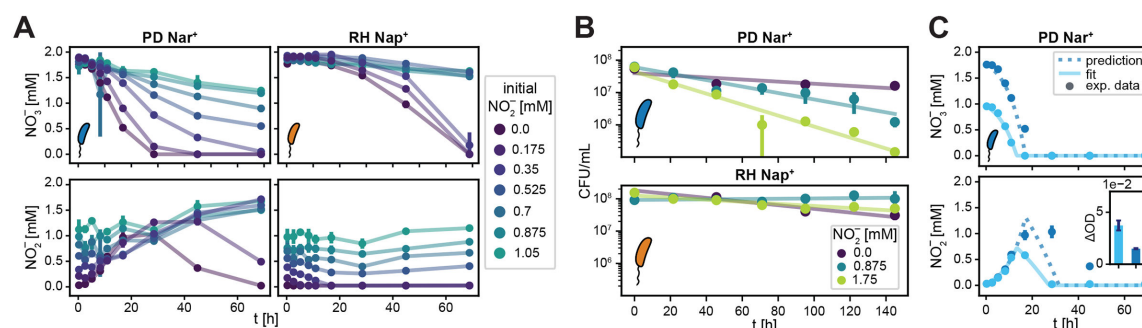


Figure 4: Nitrite toxicity impacts denitrification activity of isolates at low pH.

(A) PD Nar^+ and RH Nap^+ were grown in monoculture under anaerobic conditions at pH 6.0 with 1.75 mM nitrate and varying nitrite levels indicated by colors shown in the legend. The dynamics of nitrate and nitrite concentrations are shown for each strain. The growth of both strains was increasingly inhibited as the initial supply of nitrite ($[\text{NO}_2^-]_0$) increased. PD Nar^+ was unable to fully reduce nitrate when $[\text{NO}_2^-]_0 > 0.35$ mM (blue and green curves, top left panel) and was unable to fully reduce nitrite when $[\text{NO}_2^-]_0 > 0$ mM (all except dark purple curve, bottom left panel). Similarly, RH Nap^+ was unable to fully reduce either nitrate or nitrite when $[\text{NO}_2^-]_0 > 0.35$ mM (blue and green curves, right panels). The mean and standard deviation of biological replicates ($n = 3$) are shown. (B) PD Nar^+ and RH Nap^+ were again grown in anaerobic monoculture at pH 6.0 with varying nitrite levels; nitrate was not supplied to prevent growth (Methods). The density of colony-forming units was measured via plating with replicates ($n = 3$) for each condition, with points indicating means across replicates. Error bars were calculated by weighting across dilution levels, as described in Methods. Lines are log-linear fits, and the (negative) slope of the line indicates the mortality rate (Methods). For PD Nar^+ , inferred death rates were 0.008 ± 0.004 , 0.023 ± 0.004 , and $0.039 \pm 0.004 \text{ h}^{-1}$ for $[\text{NO}_2^-]_0 = 0$, 0.875, and 1.75 mM, respectively. For RH Nap^+ , inferred death rates were 0.013 ± 0.004 , -0.001 ± 0.001 , and $0.007 \pm 0.001 \text{ h}^{-1}$ for $[\text{NO}_2^-]_0 = 0$, 0.875, and 1.75 mM, respectively. Uncertainties indicate standard deviations of log-linear fit parameters. (C) Nitrate (top) and nitrite (bottom) metabolite dynamics for monocultures of PD Nar^+ , supplied with initial nitrate concentrations of 0.875 mM (light blue) and 1.75 mM (dark blue). Inset in the bottom panel shows the change in optical density during the 72 h growth cycle ($\Delta\text{OD} = \text{OD}_t - \text{OD}_0$); endpoint biomass levels were greater in the 0.875 mM nitrate condition than in the 1.75 mM condition, suggesting mortality induced by nitrite accumulation. A consumer-resource model was fit to the low nitrate condition (solid light blue lines) and used to predict high nitrate condition (dashed dark blue lines). The mean and standard deviation of biological replicates ($n = 3$) of the data are shown.

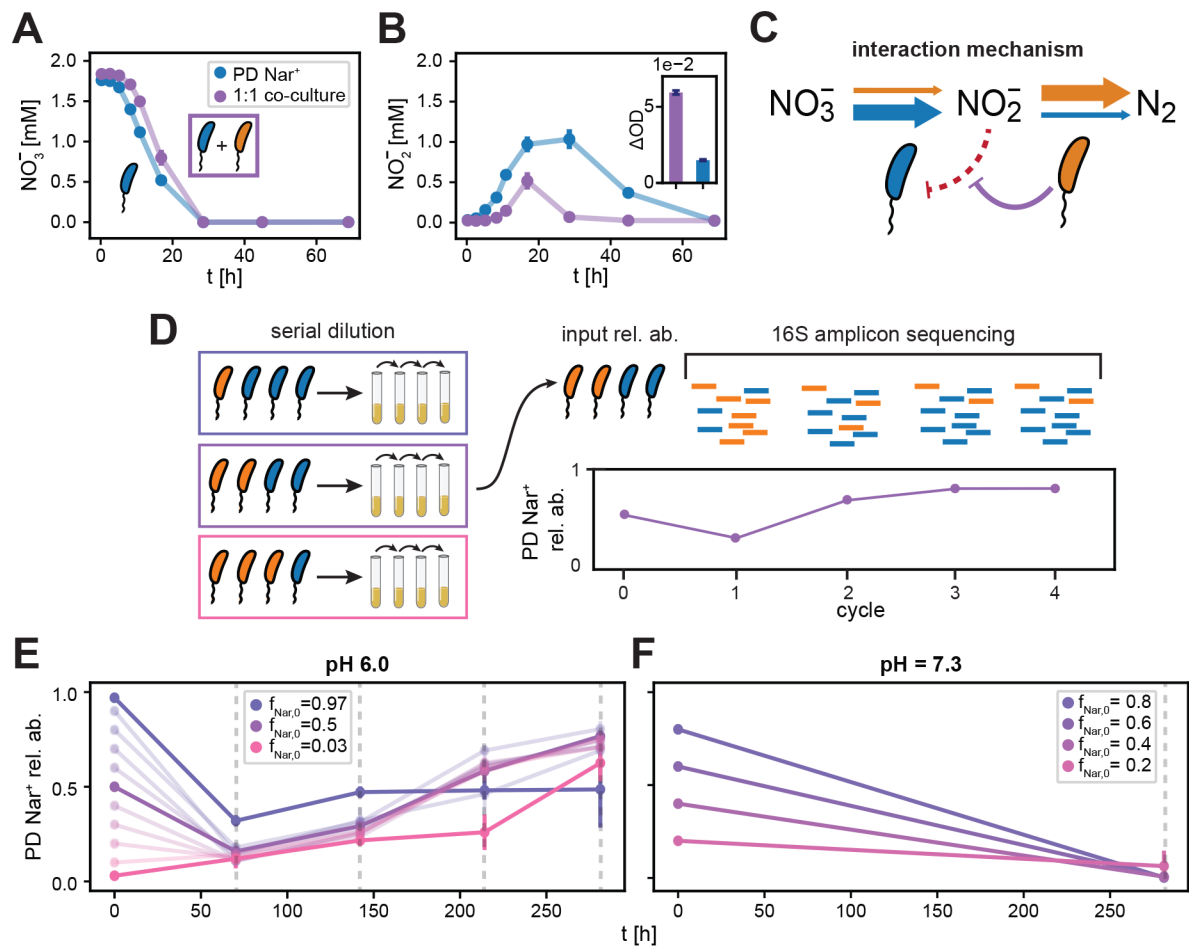


Figure 5: Coculture alleviates nitrite toxicity under acidic conditions.

(A–B) Monoculture PD Nar⁺ growth metabolite dynamics (shown previously in Fig. 4C) compared with the those of a 1:1 PD Nar⁺ and RH Nap⁺ co-culture at pH 6.0. Panel A shows nitrate concentration and panel B shows nitrite concentration. Initial biomass (OD₆₀₀ = 0.01) and nitrate (1.75 mM) was the same in each condition, but the co-culture (purple) exhibited significantly less nitrite accumulation despite only slightly decreased nitrate reduction rate, as well as significantly higher biomass production (purple bar, inset, panel B). Points and error bars indicate the means and standard deviations, respectively, across biological replicates (*n* = 3). (C) Schematic of proposed interaction between PD Nar⁺ and RH Nap⁺. (D) Schematic illustrating the multi-cycle co-culture experiment shown in panels E and F. Mixtures of PD Nar⁺ and RH Nap⁺ were prepared across a range of ratios spanning 0.03:0.97 to 0.97:0.03 (distinguished by color), with total biomass held constant. These mixtures were then transferred to fresh media buffered at pH 6.0 or pH 7.3, with 2 mM nitrate supplied. Cultures were grown under anaerobic conditions for 72 h and passaged 1:8 into fresh media for a total of four growth cycles. 16S amplicon sequencing was used to infer PD Nar⁺ relative abundance at the end of each cycle in pH 6.0 conditions, and at the end of four cycles in pH 7.3 (Fig. S7, Methods). (E) PD Nar⁺ relative abundance dynamics at pH 6.0 are shown. $f_{\text{Nar}^+,0} = 0.03, 0.5, \text{ and } 0.97$ are highlighted by darker lines. (F) PD Nar⁺ relative abundance dynamics at pH 7.3 are shown. Relative abundance values in panels

E and F are the means of inferred relative abundances across biological replicates ($n = 4$ for PD Nar⁺ relative abundance ≥ 0.5 at pH 6 and relative abundance = 0.03 in both pH conditions; $n = 3$ for all other conditions). Errorbars are calculated by adding the errors due to inference of amplification bias during sequencing and variance between biological replicates in quadrature (see Methods for detail).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

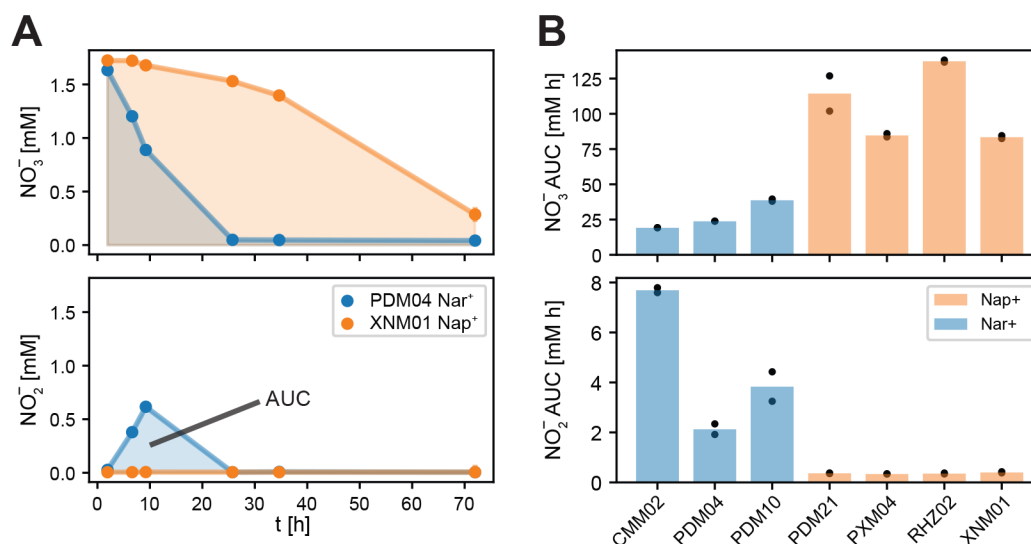


Figure 6: Nar⁺ and Nap⁺ phenotypes are conserved across diverse taxa.

Additional Nar⁺ and Nap⁺ soil isolates were grown in anaerobic conditions at pH 6.0, and their denitrification phenotypes were measured. **(A)** Denitrification phenotypes were summarized by area under the nitrate (top) and nitrite (bottom) curves (AUC). Metabolite dynamics for PDM04 Nar⁺ (blue) and XNM01 Nap⁺ (orange) monocultures are shown, with area under the curve (AUC) for each metabolite illustrated by the colored shaded regions. Smaller AUC indicates faster metabolite reduction. **(B)** AUC for nitrate (top) and nitrite (bottom) for three Nar⁺ and four Nap⁺ strains isolated in Ref. 31. Nar⁺ strains exhibit faster nitrate reduction (blue bars, top panel) and slower nitrite reduction (blue bars, bottom panel), and Nap⁺ strains are the opposite. The height of the bar corresponds to the mean AUC across biological replicates ($n = 2$).