nature microbiology

Article

https://doi.org/10.1038/s41564-024-01752-4

Environmentally dependent interactions shape patterns in gene content across natural microbiomes

Received: 14 November 2023

Accepted: 3 June 2024

Published online: 08 July 2024

Check for updates

Kyle Crocker[1,2,3], Kiseok Keith Lee[1,2,3], Milena Chakraverti-Wuerthwein[2,3,4], Zeqian Li[1,2,5], Mikhail Tikhonov[6], Madhav Mani[7,8,9,10], Karna Gowda[11,12] & Seppe Kuehn[1,2,3,10]

Sequencing surveys of microbial communities in hosts, oceans and soils have revealed ubiquitous patterns linking community composition to environmental conditions. While metabolic 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 road map for dissecting how associations between environmental variables and gene abundances arise from environmentally modulated community interactions.

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 feedback,

[1]Department of Ecology and Evolution, The University of Chicago, Chicago, IL, USA. [2]Center for the Physics of Evolving Systems, The University of Chicago, Chicago, IL, USA. [3]Center for Living Systems, The University of Chicago, Chicago, IL, USA. [4]Biophysical Sciences Graduate Program, The University of Chicago, Chicago, IL, USA. [5]Department of Physics, The University of Illinois at Urbana-Champaign, Urbana, IL, USA. [6]Department of Physics, Washington University in St. Louis, St. Louis, MO, USA. [7]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL, USA. [8]Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA. [9]NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL, USA. [10]National Institute for Theory and Mathematics in Biology, Northwestern University and The University of Chicago, Chicago, IL, USA. [11]Department of Microbiology, The Ohio State University, Columbus, OH, USA. [12]Biophysics Graduate Program, The Ohio State University, Columbus, OH, USA. e-mail: karna.gowda@gmail.com; seppe.kuehn@gmail.com

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 show 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.

In this Article, 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 (MAGs), 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[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 Text 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 and 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 (Supplementary Table 1) and inoculated them into chemically defined media buffered at pH 6.0 and pH 7.3. We then passaged these communities serially for 12 growth cycles under anaerobic conditions for 72 h each, with eightfold dilutions between cycles, supplying 1.75 mM nitrate at the beginning of each cycle (Fig. 2a and Methods). We chose these conditions because pH in soils is often strongly buffered[37–39]; millimolar 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).

End-point 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%.

https://doi.org/10.1038/s41564-024-01752-4

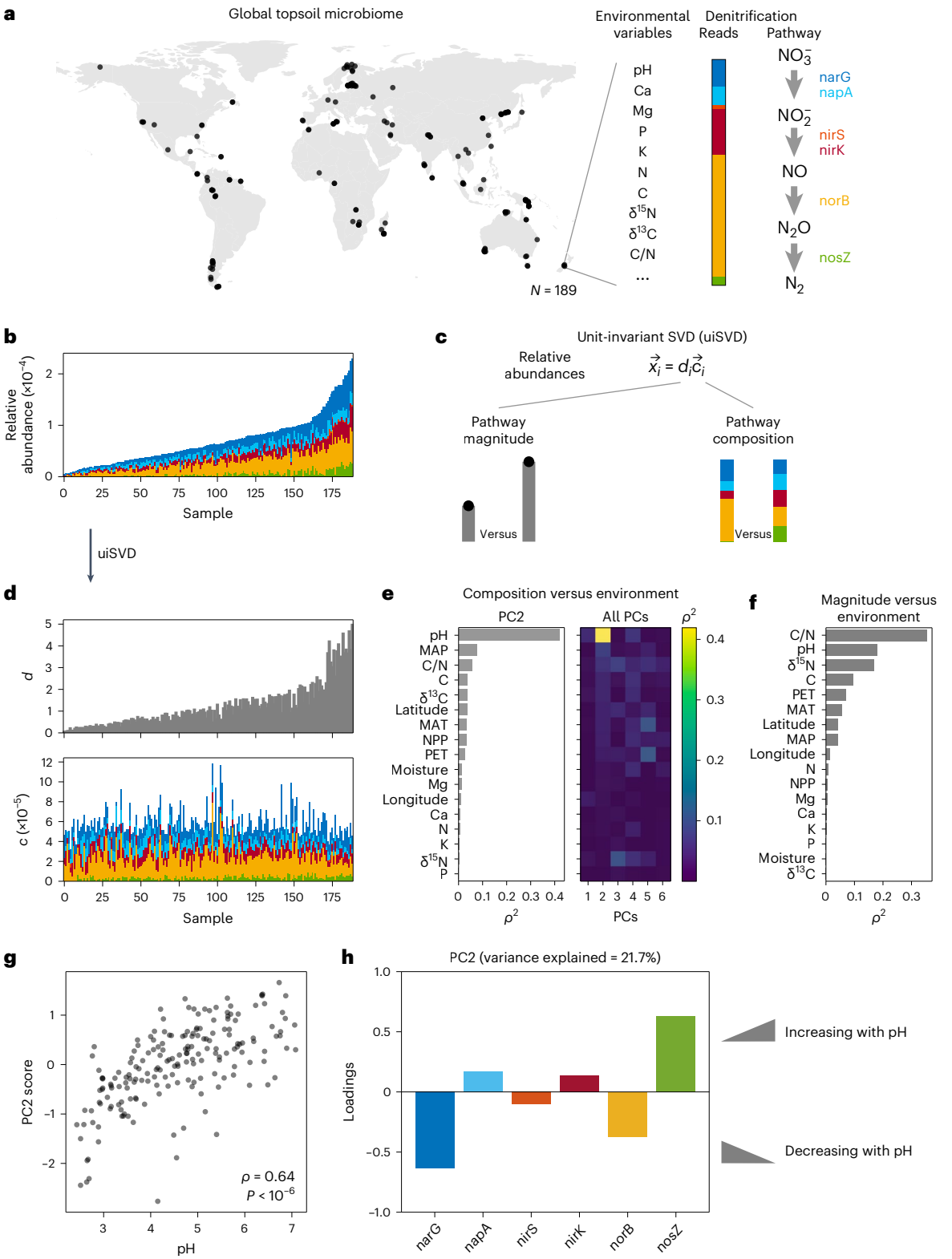


Fig. 1 | pH is associated with covariation in denitrification pathway composition in the global topsoil microbiome. a, Topsoils sampled at $n = 189$ globally distributed sites were chemically characterized (pH, Ca, Mg, ...), sequenced via shotgun metagenomics and functionally annotated by Bahram et al.[4]. **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 colour legend indicated in **a**. **c**, uiSVD[32] was used to decompose the data in **b** into contributions due to pathway magnitude (d_i) and pathway composition ($\vec{c}_i$). **d**, The results of the decomposition in **c** are plotted in **d**. **e**,**f**, PC scores for pathway composition (**e**) and pathway magnitudes (**f**) obtained via uiSVD (Methods) are compared with 17 environmental variables, and squared Pearson correlation coefficients are shown. MAP, mean annual precipitation; MAT, mean annual temperature; NPP, net primary productivity; PET, potential evapotranspiration. **g**, Scores of PC2 are most correlated with pH ($\rho = 0.64$, $P < 10^{-6}$ via one-tailed randomization test), 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 and 2.

https://doi.org/10.1038/s41564-024-01752-4

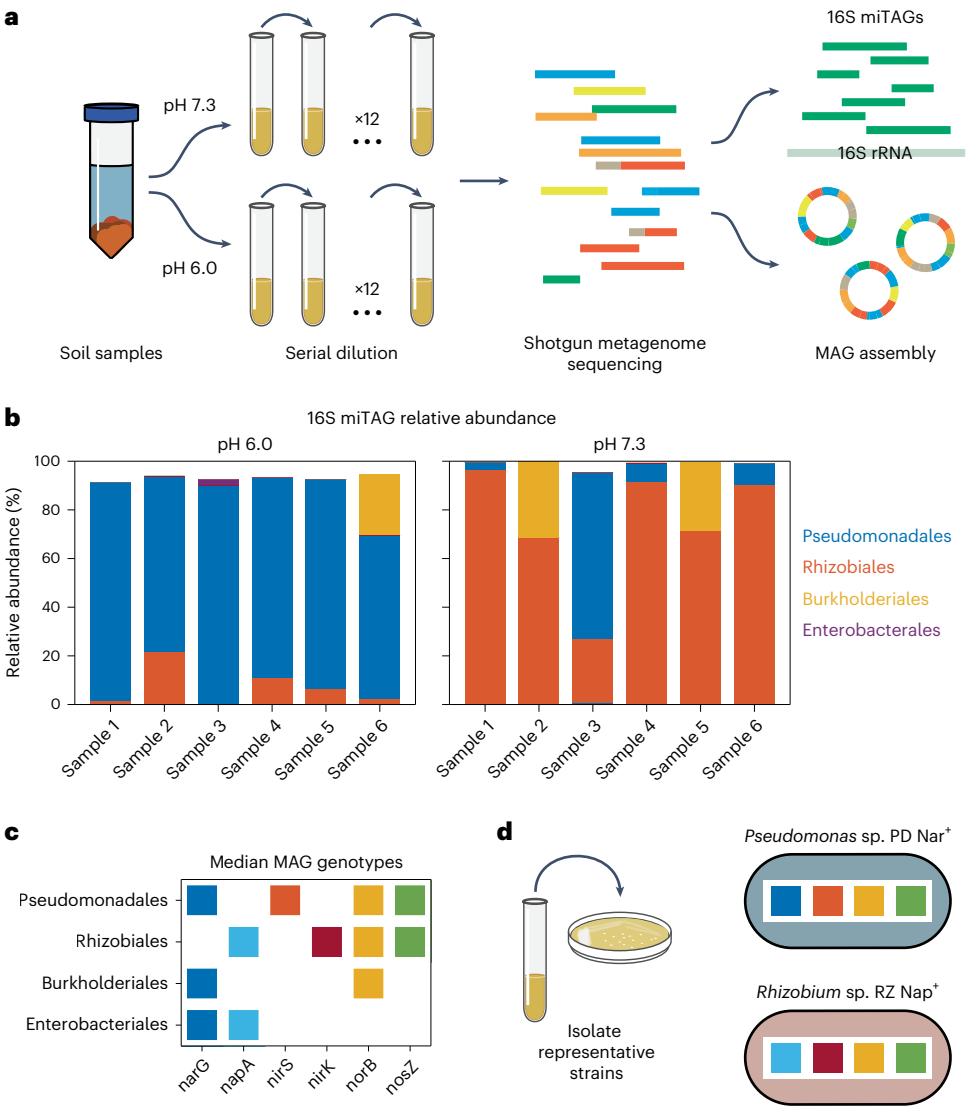


Fig. 2 | Enrichment cultures reproduce patterns in denitrification gene content in the topsoil microbiome. a, Denitrifying communities were enriched from soil samples in two pH conditions (pH 6.0 and pH 7.3) and then sequenced via shotgun metagenomics to measure taxonomic composition and genotypes. Six 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 times) into a defined medium containing 1.75 mM nitrate via a 1/8 dilution factor (Methods). **b**, End-point community compositions are shown at the level of taxonomic order for enrichments in pH 6.0 and pH 7.3. Compositions were determined by extracting 16S rRNA fragments (miTAGs) from shotgun metagenome data[72] and then taxonomically annotating miTAGs via RDP[75]. Taxa present at a relative abundance less than 1% are omitted. **c**, Median denitrification reductase genotypes are shown for MAGs corresponding to the four most abundant taxa present in the composition data in **b**. MAGs were functionally annotated using RAST[81], and the median genotype was computed over MAGs obtained in different samples. **d**, Strains were isolated from cryopreserved samples from the enrichment end point (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 Supplementary Fig. 6 and Extended Data Fig. 3.

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 pH 5.5 (Extended Data Fig. 3 and 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 Fig. 2b,c), and alternatively a greater prevalence of *napA* and *nirK* at neutral pH (Rhizobiales in Fig. 2b,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 end point from the genera *Pseudomonas* (*narG*/*nirS*) and *Rhizobium* (*napA*/*nirK*) (Fig. 2d and 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

https://doi.org/10.1038/s41564-024-01752-4

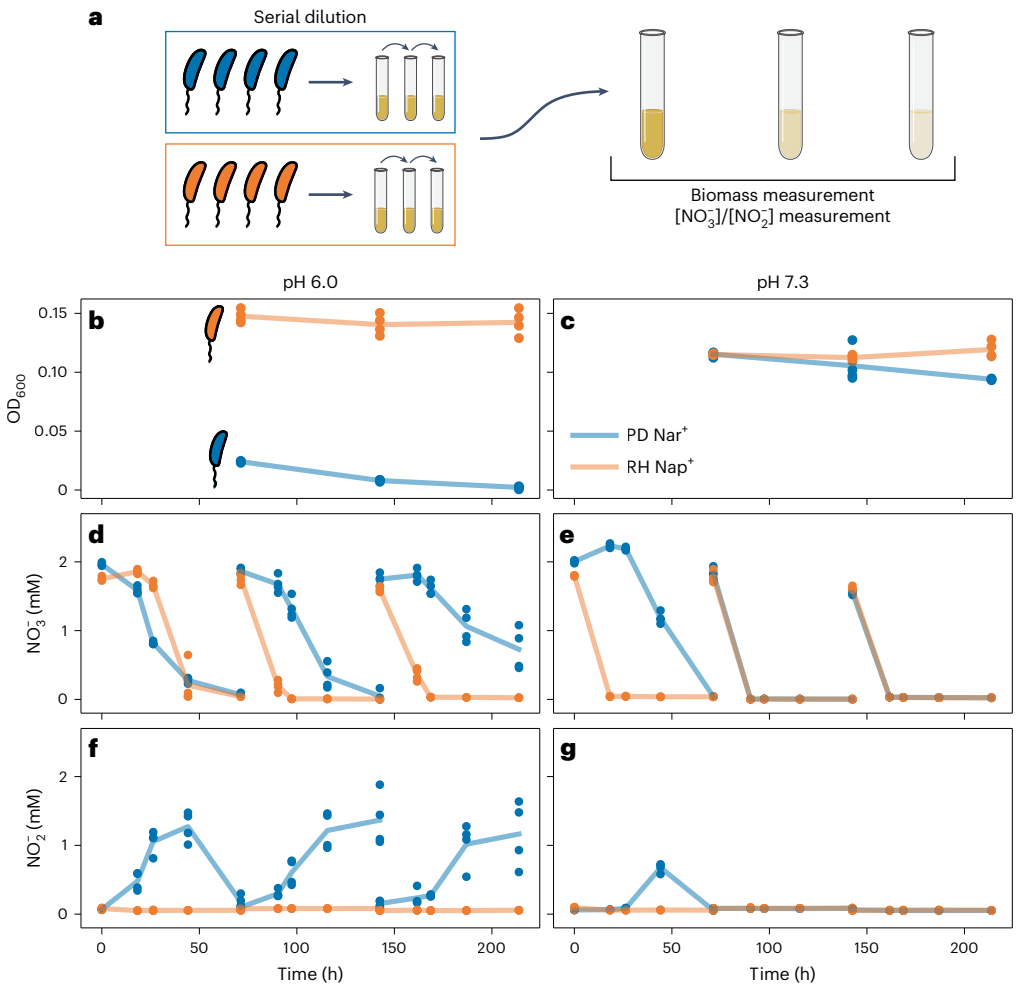


Fig. 3 | Individual traits do not explain the outcome of acidic enrichments. **a**, Schematic of the experimental design. PD Nar$^+$ (blue) and RH Nap$^+$ (orange) were passaged repeatedly in monoculture under denitrifying (anaerobic) conditions. After each 72 h growth cycle, cultures were passaged (3 times) 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**,**c**, End-point biomass is shown for each strain at both pH 6.0 (**b**) and pH 7.3 (**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 (**d** and **e**) and nitrite (**f** and **g**) concentration dynamics are shown at six time points 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 (**d** and **f**, blue). PD Nar$^+$ (blue) accumulates nitrite at each pH condition, whereas RH Nap$^+$ (orange) does not (**f** and **g**). Biological replicates ($n = 4$) are shown for each strain in each experimental condition, with lines connecting the averages of these replicates.

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). In addition, simulated competition using a consumer-resource model previously developed for denitrification[31] confirmed the fitness of RH Nap$^+$ over PD Nar$^+$ at pH 7.3 based on monoculture phenotypes (Supplementary Text Section 2.1 and Extended Data Fig. 4).

By 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). Furthermore, the nitrate reduction rate of PD Nar$^+$ slowed gradually with increasing initial nitrite concentration, and the nitrate reduction rate for RH Nap$^+$ showed a threshold-like dependence on nitrite concentration ($[NO_2^-] \approx 0.4$–0.5 mM).

https://doi.org/10.1038/s41564-024-01752-4

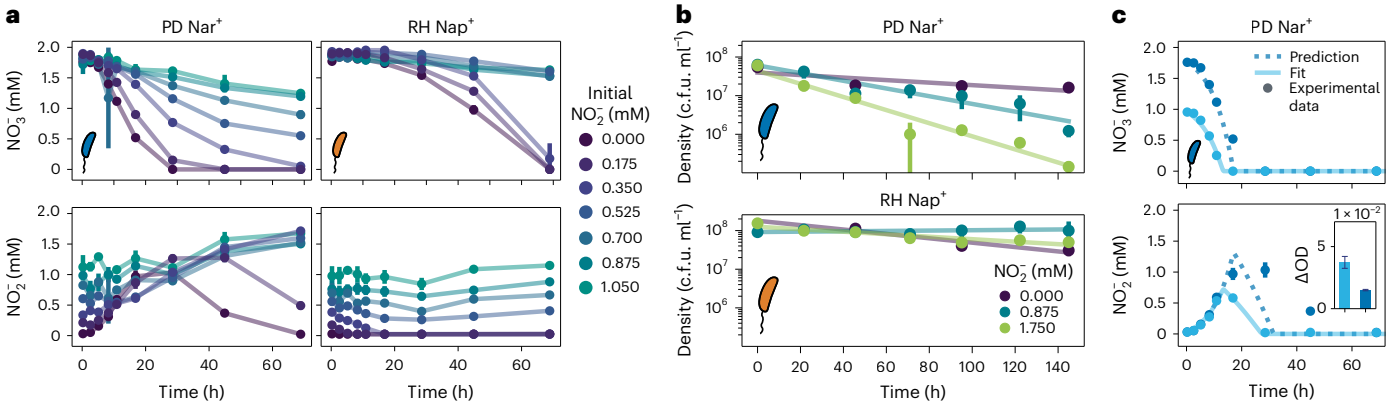


Fig. 4 | Nitrite toxicity impacts denitrification activity of isolates at low pH. **a**, PD Nar$^+$ and RH Nap$^+$ were grown in monoculture under anaerobic conditions at pH 6.0 with 1.75 mM nitrate and varying nitrite levels indicated by colours shown in the legend. The dynamics of nitrate and nitrite concentrations are shown for each strain. The growth of both strains was increasingly inhibited as the initial supply of nitrite ($[NO_2^-]_0$) increased. PD Nar$^+$ was unable to fully reduce nitrate when $[NO_2^-]_0 > 0.35$ mM (blue and green curves, top left panel) and was unable to fully reduce nitrite when $[NO_2^-]_0 > 0$ mM (all except dark purple curve, bottom left panel). Similarly, RH Nap$^+$ was unable to fully reduce either nitrate or nitrite when $[NO_2^-]_0 > 0.35$ 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 h^{-1}, 0.023 ± 0.004 h^{-1} and 0.039 ± 0.004 h^{-1} for $[NO_2^-]_0 = 0 \sim$ mM, 0.875 mM and 1.75 mM, respectively. For RH Nap$^+$, inferred death rates were 0.013 ± 0.004 h^{-1}, -0.001 ± 0.001 h^{-1} and 0.007 ± 0.001 h^{-1} for $[NO_2^-]_0 = 0 \sim$ mM, 0.875 mM 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 (OD) during the 72 h growth cycle ($\Delta OD = OD_{final} - OD_{initial}$); end-point 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.

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 and 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 and Supplementary Text Section 2.2). Overall, these experiments showed that nitrite accumulation has a substantial 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 (Figs. 3f and 4a), 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 ~0.5 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 and 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 and 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 show nitrate and nitrite specialist phenotypes, respectively.

We phenotyped seven strains isolated in our previous study[31] which span α-, β- and γ-proteobacteria and possess one or the other nitrate reductase (three Nar$^+$, four Nap$^+$). To avoid ambiguity, we excluded strains possessing both Nar and Nap. These reductases tend to appear separately in reference genomes (Supplementary Text Section 3

https://doi.org/10.1038/s41564-024-01752-4

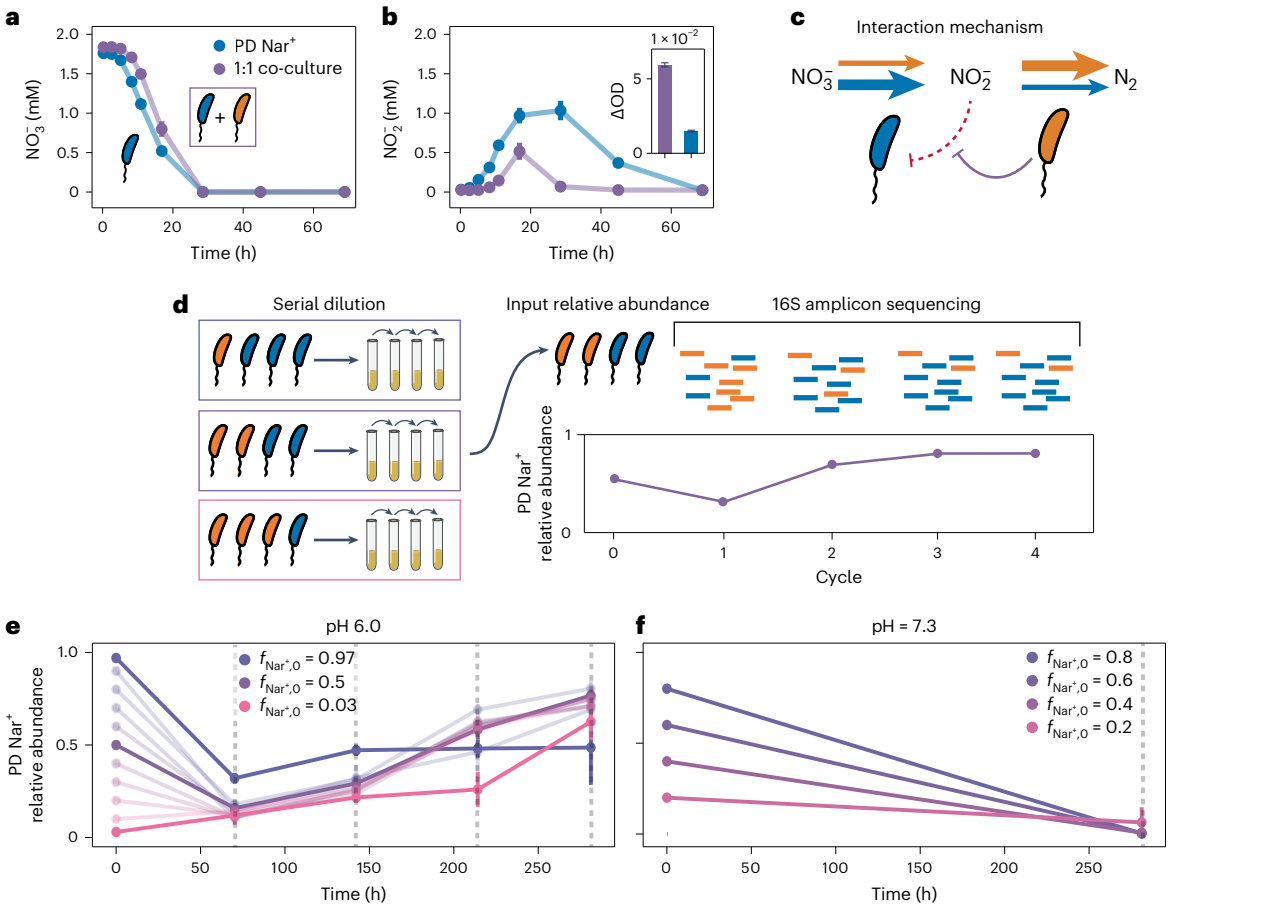


Fig. 5 | Co-culture alleviates nitrite toxicity under acidic conditions. **a**,**b**, Monoculture PD Nar$^+$ growth metabolite dynamics (shown previously in Fig. 4c) compared with those of a 1:1 PD Nar$^+$ and RH Nap$^+$ co-culture at pH 6.0. **a**, Nitrate concentration. **b**, Nitrite concentration. Initial biomass (OD_{600} = 0.01) and nitrate (1.75 mM) were the same in each condition, but the co-culture (purple) showed significantly less nitrite accumulation despite only slightly decreased nitrate reduction rate, as well as significantly higher biomass production (purple bar, inset, **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 **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 colour), 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 (Methods and Supplementary Fig. 7). **e**, PD Nar$^+$ relative abundance dynamics at pH 6.0. $f_{Nar^+,0}$ = 0.03, 0.5 and 0.97 are highlighted by darker lines. **f**, PD Nar$^+$ relative abundance dynamics at pH 7.3. Relative abundance values in **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). Error bars are calculated by adding the errors due to inference of amplification bias during sequencing and variance between biological replicates in quadrature (see Methods for details).

and Supplementary Fig. 1), and when they co-occur Nap may be used for redox balancing[47] or expressed primarily in micromolar 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 showed 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 harbour 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 Text Section 4). Briefly, we found that (a) a substantial fraction of denitrification reductase reads can be classified as Proteobacterial (Supplementary Fig. 2) and (b) laboratory incubation experiments performed on homogenized topsoils under anaerobic conditions show Protobacterial populations dominate community response to nitrate amendment (Supplementary Fig. 3). Finally, a meta-analysis of culture-based studies of denitrifiers in soils also supports the claim that Proteobacteria dominate these populations (Supplementary Table 2). 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

https://doi.org/10.1038/s41564-024-01752-4

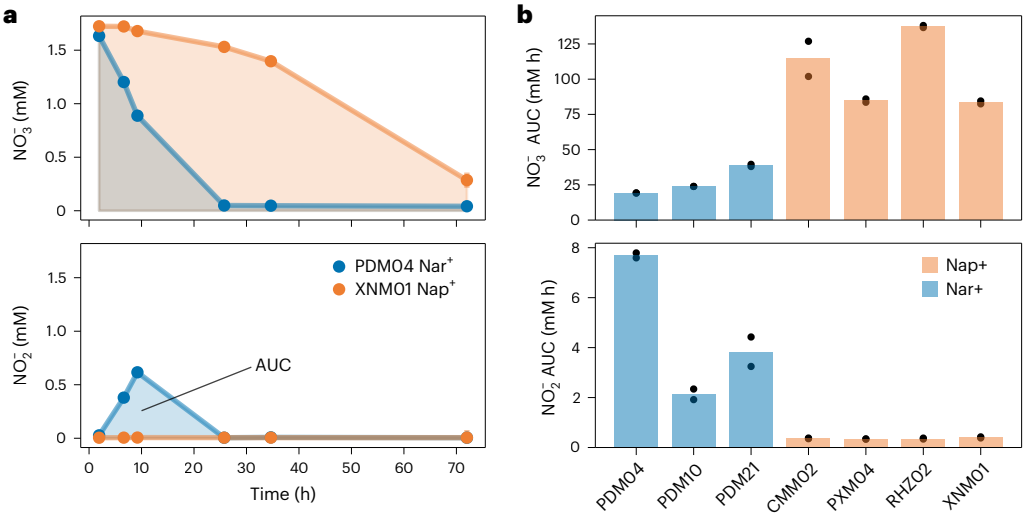


Fig. 6 | Nar$^+$ and Nap$^+$ phenotypes are conserved across diverse taxa. Additional Nar$^+$ and Nap$^+$ soil isolates were grown in anaerobic conditions at pH 6.0, and their denitrification phenotypes were measured. **a**, Denitrification phenotypes were summarized by area under the nitrate (top) and nitrite (bottom) curves (AUC). Metabolite dynamics for monocultures of isolates PDM04 Nar$^+$ (blue) and XNM01 Nap$^+$ (orange) are shown, with AUC for each metabolite illustrated by the coloured 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 show 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 ($n = 2$).

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 (Supplementary Fig. 3 and Supplementary Text Section 4). Specifically, a small fraction of all taxa in the soil community respond to the addition of nitrate (0.3 ± 0.2%; Supplementary Fig. 3c). The result suggests that despite the ecological complexity of soils, a small number of taxa (4 ± 3 in our experiment; Supplementary Fig. 3c) 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 (Supplementary Fig. 1), and previous work suggests that Nap may be used for redox balancing[47] or micromolar nitrate concentrations[48,49] in strains that contain both enzymes. In addition, 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[31]. However, this work also shows that strains with *narG* exhibit faster nitrate reduction than strains with only *napA*[31], a finding corroborated by in vitro comparisons of these enzymes[65–69]. 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[35,70]. 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 (Supplementary 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 (Supplementary Figs. 4 and 5 and Supplementary Text 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 analysed 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, ^{12}C, ^{13}C, ^{14}N and ^{15}N. In addition, site data including mean annual temperature, mean annual precipitation, potential evapotranspiration, net primary productivity and moisture were obtained from public databases. Paired metagenome reads were quality filtered and annotated to yield Kyoto Encyclopedia of Genes and Genomes (KEGG)

orthologue 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 KEGG orthologues 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 KEGG orthologues were computed by dividing the number of reads mapped to each KEGG orthologue 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, δ^{13}C, δ^{15}N, C/N, latitude, longitude, mean annual temperature, mean annual precipitation, potential evapotranspiration, net primary productivity and moisture.

uiSVD. uiSVD[32] was applied to the denitrification reductase relative abundance matrix (X) to separate variation in pathway magnitude (that is, variations in total reductase content) from variation in pathway composition (that is, variations in relative fractions of reductase content). This approach was motivated by modelling the data as

$$\vec{x}_i = d_i \vec{c}_i, \tag{1}$$

where $\vec{x}_i$ is the ith row of X (with entries being relative abundances of each KEGG orthologue), 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

$$X = DC, \tag{2}$$

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

$$X = D\tilde{U}\tilde{S}\tilde{V}^{\mathrm{T}}E, \tag{3}$$

where D, $\tilde{S}$ and E are diagonal matrices, $\tilde{U}$ and $\tilde{V}$ are unitary matrices, and T denotes the matrix transpose. Hence,

$$C = \tilde{U}\tilde{S}\tilde{V}^{\mathrm{T}}E. \tag{4}$$

The results of this decomposition are shown in Fig. 1d. Note that, in this representation, D encodes the scale of the rows of X (that is, pathway magnitude), while E encodes the relative scales of the columns (that is, typical relative scales of each gene). Code used to perform uiSVD was adapted for Python from the MATLAB code given in ref. 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. 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, principal components analysis was applied to $\tilde{C}$ by mean-centring the columns and applying singular value decomposition to the resulting matrix, yielding a decomposition USV^T. The columns of V represent the orthonormal PCs of $\tilde{C}$ (shown in Extended Data Fig. 1), and

$$T = US \tag{5}$$

represents the 'scores' matrix, that is, the projections of $\tilde{C}$ onto the PCs in V.

Analysing gene–environment covariation. To relate gene covariation to variation in site environmental variables, Pearson correlations (ρ^2) were computed between the scores of each PC (equation (5)) and each environmental variable (squared Pearson correlations shown in Fig. 1e). In addition, 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 (not a number) 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 version 0.39 with default settings. Trimmed reads were assembled using SPAades version 3.15.0. Predicted open reading frames on assembled contigs were then annotated as KEGG orthologue groups of the denitrification reductases (*narG*/K00370, *napA*/K02567, *nirS*/K15864, *nirK*/K00368, *norB*/K04561, *nosZ*/K00376) using eggNOG-mapper version 2.0.8, and reads were mapped to these annotated open reading frames using minimap2. Finally, reads mapping to denitrification KEGG orthologues were phylogenetically classified using the Kaiju web server (https://kaiju.binf.ku.dk/server; ref. 71). The results of these classifications, at the level of taxonomic phylum and grouping all samples together, are shown in Supplementary Fig. 2.

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 Supplementary Table 1. 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 2 months before the start of the experiment. No sample compositing was performed. To mechanically homogenize soils before 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 the 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 pH 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 µl 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% N_2/CO_2 mixture. The plate was incubated at 30 °C and shaken at 950 r.p.m. (Talboys Professional 1000MP, 3 mm orbital radius).

Cultures were grown under these conditions for 72 h. At the end of this time, 150 µl of the cultures was passaged under anaerobic conditions into a freshly prepared plate containing 1,050 µl 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 end-point nitrate and nitrite concentrations via Griess assay and vanadium (III) chloride

reduction method via the protocol described in ref. 31 (Supplementary Fig. 6). In addition, optical densities at 600 nm (OD_{600}) were also recorded (BMG CLARIOstar) using 300 µl of end-point cultures in 96-well optical plates (Supplementary Fig. 6).

Cultures were repeatedly passaged and assayed in this manner for 12 cycles. At the end points of cycles 4, 8 and 12, 100 µl samples of end-point cultures were cryopreserved by adding 100 µl 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 end-point (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 × 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 ribosomal RNA fragments (miTAGs) were extracted from the sequencing data using the miTAGs extraction script version 1[72], after trimming and merging of overlapping paired-end reads via Trim-Galore version 0.6.7[73] and BBMerge version 38.22[74], respectively. The resulting miTAGs were then taxonomically classified using the RDP (Ribosomal Database Project) Classifier[75], using a confidence threshold of 80% and the copy number adjustment option enabled.

The sequencing data were then binned into MAGs using the metaWRAP pipeline version 1.3[42]. In summary, the pipeline performed adapter and quality trimming of raw reads using Trim-Galore version 0.6.7[73]; assembly using metaSPAdes version 3.15.4[76]; binning using CONCOCT version 1.1.0[77], MaxBin2 version 2.2.7[78] and metaBAT version 2.15[79]; bin quality assessment using CheckM version 1.0.11[80]; 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 (Supplementary Table 3). High-quality MAGs were then annotated using the RAST server[81].

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 number 1. Glycerol-cryopreserved cultures were streaked to purity on 1/10× TSB agar plates (1.5% agar *w*/*v*) in aerobic conditions. Overnight cultures of single-colony isolates were grown in 1/10× TSB (30 °C, 400 r.p.m.) 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[82–84]. 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 pH 7.3. Seven soil samples (soil numbers 7–13; Supplementary Table 1) were taken from prairies and forest preserves across the Midwestern United States (Illinois, Indiana, Michigan and Wisconsin) 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: pH 5.0, pH 5.5, pH 6.0 and pH 7.3. In pH 5.0 and pH 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 and sequencing and data analysis of cycle 12 cultures are as described above, except that defined media succinate concentration is reduced to 4 mM. 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 end-point enrichment are shown in Extended Data Fig. 3.

Characterization of isolate phenotypes

Culturing protocol. Strains were pre-cultured in two stages under aerobic conditions before 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 r.c.f. (relative centrifugal force) (400 r.p.m. in Fisherbrand Incubating Microplate Shakers 02-217-759, 3 mm orbital radius or 219 r.p.m. 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 1.7 ml of SDM at pH 7.3 with 25 mM succinate (and no sodium nitrate). Wells were then inoculated with 17 µl of the saturated R2B PD Nar$^+$ and RH Nap$^+$. After sealing, the cultures were incubated at 0.5 r.c.f. and 30 °C in aerobic conditions overnight. These cultures reached saturation during this time. Saturated SDM culture densities were measured and normalized to OD_{600} = 1.0 via dilution into pH 7.4 phosphate-buffered saline (8 g l^{-1} H_2O, 0.2 g l^{-1} KCl, 2.68 g l^{-1} $Na_2HPO_4{\cdot}7H_2O$, 0.24 g l^{-1} KH_2PO_4).

Due to a lag time associated with growth in anaerobic conditions for facultative anaerobes[85–90], 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 (4 mM 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 µl of optical density-normalized PD Nar$^+$ and RH Nap$^+$ aerobic pre-cultures, resulting in a starting optical density 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 r.p.m. (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 end-point anaerobic pre-cultures were measured using 300 µl of cultures in 96-well optical plates. Optical densities were then normalized to 0.04 via dilution into SDM (no nitrate). About 150 µl of this normalized end-point anaerobic culture was then passaged to 1,050 µl of SDM, with pH and nitrate/nitrite concentrations varying according to the experiment (for example, Fig. 4). In the case of co-culture phenotyping experiments, this 150 µl 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 r.p.m. 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. End-point biomass was measured via OD_{600} using 300 µl samples.

Additional experiments were conducted where measurements were taken over multiple cycles of growth and dilution into fresh

medium (for example, Figs. 3 and 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. About 12 µl of normalized culture from the end point 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 end-point 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[91]. 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 and 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 s at 95 °C, 30 s at 55 °C, 30 s 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[91]. 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[92] and then quality filtered with DADA2 plugin in the QIIME2 pipeline with default parameters[93]. Taxonomy of amplicon sequence variants (ASVs) was assigned by q2-feature-classifier prefitted to the SILVA database using the naive Bayes algorithm for the V4 region of 16S rRNA[94]. ASVs of the family *Rhizobiaceae* were classified as RH Nap$^+$, whereas 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[95]. 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. 95, we assumed the biomass x_* of each strain is proportional to the ASV counts C_*, with bias factors b_*:

$$\begin{aligned} x_{\mathrm{PD}} &= b_{\mathrm{PD}} C_{\mathrm{PD}} \\ x_{\mathrm{RH}} &= b_{\mathrm{RH}} C_{\mathrm{RH}} \end{aligned} \tag{6}$$

The relative abundance then becomes

$$f_{\mathrm{PD}} = \frac{C_{\mathrm{PD}}}{C_{\mathrm{PD}} + b_{\mathrm{r}} C_{\mathrm{RH}}}. \tag{7}$$

where b_{r} is the ratio of the bias factors $b_{\mathrm{r}} = b_{\mathrm{RH}}/b_{\mathrm{PD}}$. Using measured counts for cultures with known PD Nar$^+$ relative abundances, we fit b_{r} using the Levenberg–Marquardt algorithm (Supplementary Fig. 7) and obtained a bias ratio of 4.3 ± 0.7. Equation (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[96]. To account for (b), standard deviation of relative abundance between biological replicates was calculated. The sum of 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 pH 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 1/10× TSB, 1.5% agar plates (1.5 g l^{-1} tryptone, 0.5 g l^{-1} soytone, 0.5 g l^{-1} sodium chloride, 15 g l^{-1} Bacto agar) and that RH Nap$^+$ colonies were most distinguishable on a defined medium with mannitol as the carbon source. Mannitol-defined medium 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 (that is, two stages of aerobic growth and one stage of anaerobic growth). After the anaerobic pre-culture step, 150 µl of culture of each strain normalized to OD_{600} = 0.04 was passaged under anaerobic conditions to 1,050 µl of SDM (4 mM succinate, no nitrate) and incubated at 30 °C and 950 r.p.m. Two pH conditions (pH 6 and pH 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. About 10 µl 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. About 10 µl 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 1/10× TSB and 2 days for RH Nap$^+$ on mannitol-defined medium).

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

$$n_{\mathrm{i}} = D_{\mathrm{i}} C_{\mathrm{i}}, \tag{8}$$

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[97]:

$$n_{\mathrm{c}} = \frac{\sum_{\mathrm{i}} \bar{n}_{\mathrm{i}} w_{\mathrm{i}}}{\sum_{\mathrm{i}} w_{\mathrm{i}}} \tag{9}$$

where $w_{\mathrm{i}} = 1/\sigma_{\mathrm{i}}^2$ are the weights and $\bar{n}_{\mathrm{i}}$ 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

$$\sigma_{\mathrm{c}} = \frac{1}{\sqrt{\sum_{\mathrm{i}} w_{\mathrm{i}}}}. \tag{10}$$

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 these conditions were discarded from analysis. At pH 6.0, growth was observed at $I_0 = 0.5$ mM (monotonic decrease in nitrite concentration; Supplementary Fig. 8), 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:

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

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

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

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 (pH 5.0–7.1) from the Cook Agronomy Farm (CAF) in September 2022 (Supplementary Table 4). The CAF (46.78° N, 117.09° W, 800 m above sea level) is a long-term agricultural research site located near Pullman, WA, 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)[98].

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^{-1}), 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) (Supplementary Table 4).

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 r.p.m., 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, end-point 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 end-point subsamples in a combined chemical and mechanical procedure using the PowerSoil DNeasy PowerSoil HTP 96 Kit (Qiagen). 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) 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 two-stage PCR workflow (Illumina). The V3–V4 region was amplified using forward primer 341-b-S-17 (CCTACGGGNGGCWGCAG) and reverse primer 785-a-A-21 (GACTACHVGGGTATCTAATCC)[99]. 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). A standardized 10-strain gDNA mixture (MSA-1000, ATCC) 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 preprocessing was performed with USEARCH version 11.0[100]. The filtered reads were then processed with DADA2 version 1.26 following the developers' recommended pipeline[93]. 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 4,466 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 1,000 counts across the 902 samples of this larger dataset.) Taxonomy was assigned by DADA2 using the SILVA database version 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 ± 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 a 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 $\sigma(c_0, c_{\mathrm{frac}}) = \sqrt{(c_{\mathrm{frac}} n)^2 + c_0^2}$ counts (Supplementary Fig. 9). In this expression, c_{frac} was estimated from moderate-abundance ASVs (>50 counts) for which the other noise term is negligible; c_0 was then determined as the value for which 67% of replicate–replicate comparisons are within $\pm\sigma(c_0, c_{\mathrm{frac}})$ 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 $c_{\mathrm{frac}} = 0.22 \pm 0.03$ 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 $z = \Phi^{-1}(1 - \alpha/2/n_{\mathrm{ASV}})$, where $\Phi^{-1}(x)$ is the inverse cumulative distribution function of the standard normal distribution, $\alpha = 0.01$, and n_{ASV} is 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), which are available on the Open Science Framework data repository for the present manuscript; for additional technical details, the reader is referred to the detailed comments in these scripts.

To determine which significantly enriched ASVs are likely to represent denitrifiers, PICRUSt2 version 2.5.2[101] 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 orthologues for each ASV was computed using the default parameters of the PICRUSt2 pipeline. An ASV was classified as a denitrifier if KEGG orthologues for any denitrification reductase (*narG*/K00370, *napA*/K02567, *nirS*/K15864, *nirK*/K00368, *norB*/K04561, *nosZ*/K00376) were inferred and if the dissimilatory nitrate reduction to ammonium (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 and reproducibility. No data were excluded from the analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data associated with this manuscript are publicly available at https://doi.org/10.17605/OSF.IO/N4J6F and by request. Raw sequence reads for soil enrichment experiments are deposited under National Center for Biotechnology Information BioProject ID PRJNA976277, and raw sequence reads for co-culture experiments are deposited under National Center for Biotechnology Information 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. 82–84; https://www.arb-silva.de) was used for taxonomic classification.

Code availability

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

References

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
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).
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. *Ann. Microbiol.* **69**, 1461–1473 (2019).
4. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).
5. Malik, A. A., Thomson, B. C., Whiteley, A. S., Bailey, M. & Griffiths, R. I. Bacterial physiological adaptations to contrasting edaphic conditions identified using landscape scale metagenomics. *MBio* **8**, e00799–17 (2017).
6. Sofi, M. H. et al. pH of drinking water influences the composition of gut microbiome and type 1 diabetes incidence. *Diabetes* **63**, 632–644 (2014).
7. Zinöcker, M. K. & Lindseth, I. A. The western diet–microbiome-host interaction and its role in metabolic disease. *Nutrients* **10**, 365 (2018).
8. Singh, R. K. et al. Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15**, 1–17 (2017).
9. Baas-Becking, L. G. M. *Geobiologie of Inleiding Tot de Milieukunde* (Van Stockkum & Zoon, 1934).
10. Dimitriu, P. A. & Grayston, S. J. Relationship between soil properties and patterns of bacterial β-diversity across reclaimed and natural boreal forest soils. *Microb. Ecol.* **59**, 563–573 (2010).
11. Rousk, J. et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **4**, 1340–1351 (2010).
12. Jones, R. T. et al. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* **3**, 442–453 (2009).
13. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).
14. Pacheco, A. R. & Segrè, D. A multidimensional perspective on microbial interactions. *FEMS Microbiol. Lett.* **366**, fnz125 (2019).
15. Gopalakrishnappa, C., Li, Z. & Kuehn, S. Environmental modulators of algae–bacteria interactions at scale. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.23.534036 (2023).
16. Pearl Mizrahi, S., Goyal, A. & Gore, J. Community interactions drive the evolution of antibiotic tolerance in bacteria. *Proc. Natl Acad. Sci. USA* **120**, e2209043119 (2023).
17. Hsu, R. H. et al. Microbial interaction network inference in microfluidic droplets. *Cell Syst.* **9**, 229–242 (2019).
18. Sherr, E. B. & Sherr, B. F. Significance of predation by protists in aquatic microbial food webs. *Antonie van Leeuwenhoek* **81**, 293–308 (2002).
19. de la Cruz Barron, M. et al. Shifts from cooperative to individual-based predation defense determine microbial predator–prey dynamics. *ISME J.* (2023).
20. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The evolution and ecology of bacterial warfare. *Curr. Biol.* **29**, R521–R537 (2019).
21. Venturelli, O. S. et al. Deciphering microbial interactions in synthetic human gut microbiome communities. *Mol. Syst. Biol.* **14**, e8157 (2018).
22. Gamble, T. N., Betlach, M. R. & Tiedje, J. M. Numerically dominant denitrifying bacteria from world soils 1. *Appl. Environ. Microbiol.* **33**, 926–939 (1977).
23. Seitzinger, S. et al. Denitrification across landscapes and waterscapes: a synthesis. *Ecol. Appl.* **16**, 2064–2090 (2006).

24. Lu, H., Chandran, K. & Stensel, D. Microbial ecology of denitrification in biological wastewater treatment. *Water Res.* **64**, 237–254 (2014).
25. Irrazábal, T., Belcheva, A., Girardin, S. E., Martin, A. & Philpott, D. J. The multifaceted role of the intestinal microbiota in colon cancer. *Mol. Cell* **54**, 309–320 (2014).
26. Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
27. Rodionov, D. A., Dubchak, I. L., Arkin, A. P., Alm, E. J. & Gelfand, M. S. Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Comput. Biol.* **1**, e55 (2005).
28. Shapleigh, J. P. The denitrifying prokaryotes. In *The Prokaryotes* (eds. Rosenberg, E. et al.) 769–792 (Springer, 2006).
29. Heylen, K. et al. Cultivation of denitrifying bacteria: optimization of isolation conditions and diversity study. *Appl. Environ. Microbiol.* **72**, 2637–2643 (2006).
30. Lycus, P. et al. Phenotypic and genotypic richness of denitrifiers revealed by a novel isolation strategy. *ISME J.* **11**, 2219–2232 (2017).
31. Gowda, K., Ping, D., Mani, M. & Kuehn, S. Genomic structure predicts metabolite dynamics in microbial communities. *Cell* **185**, 530–546 (2022).
32. Uhlmann, J. A generalized matrix inverse that is consistent with respect to diagonal transformations. *SIAM J. Matrix Anal. Appl.* https://doi.org/10.1137/17M113890X (2018).
33. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
34. Šimek, M. & Cooper, J. E. The influence of soil pH on denitrification: progress towards the understanding of this interaction over the last 50 years. *Eur. J. Soil Sci.* **53**, 345–354 (2002).
35. Lilja, E. E. & Johnson, D. R. Segregating metabolic processes into different microbial cells accelerates the consumption of inhibitory substrates. *ISME J.* **10**, 1568–1578 (2016).
36. Tiedje, J. M., Sexstone, A. J., Myrold, D. D. & Robinson, J. A. Denitrification: ecological niches, competition and survival. *Antonie van Leeuwenhoek* **48**, 569–583 (1983).
37. Magdoff, F. & Bartlett, R. Soil pH buffering revisited. *Soil Sci. Soc. Am. J.* **49**, 145–148 (1985).
38. Nelson, P. N. & Su, N. Soil pH buffering capacity: a descriptive function and its application to some acidic tropical soils. *Soil Res.* **48**, 201–207 (2010).
39. Bache, B. The role of calcium in buffering soils. *Plant Cell Environ.* **7**, 391–395 (1984).
40. Kreitler, C. W. & Jones, D. C. Natural soil nitrate: the cause of the nitrate contamination of ground water in Runnels County, Texas[a]. *Groundwater* **13**, 53–62 (1975).
41. Dowdell, R. J. & Cannell, R. Q. Effect of ploughing and direct drilling on soil nitrate content. *Eur. J. Soil Sci.* 26, 10.1111/j.1365-2389.1975.tb01929.x (1975).
42. Uritskiy, G. V., Diruggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 1–13 (2018).
43. Bollag, J.-M. & Henninger, N. M. Effects of nitrite toxicity on soil bacteria under aerobic and anaerobic conditions. *Soil Biol. Biochem.* **10**, 377–381 (1978).
44. Tarr, H. The action of nitrites on bacteria. *J. Fish. Board Can.* **5**, 265–275 (1941).
45. Bancroft, K., Grant, I. F. & Alexander, M. Toxicity of NO_2^-: effect of nitrite on microbial activity in an acid soil. *Appl. Environ. Microbiol.* **38**, 940–944 (1979).
46. Durvasula, K. et al. Effect of periplasmic nitrate reductase on diauxic lag of *Paracoccus pantotrophus*. *Biotechnol. Prog.* **25**, 973–979 (2009).
47. Richardson, D. J. & Ferguson, S. J. The influence of carbon substrate on the activity of the periplasmic nitrate reductase in aerobically grown *Thiosphaera pantotropha*. *Arch. Microbiol.* **157**, 535–537 (1992).
48. Wang, H., Tseng, C. P. & Gunsalus, R. P. The napF and narG nitrate reductase operons in *Escherichia coli* are differentially expressed in response to submicromolar concentrations of nitrate but not nitrite. *J. Bacteriol.* **181**, 5303–5308 (1999).
49. Potter, L. C., Millington, P., Griffiths, L., Thomas, G. H. & Cole, J. A. 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? *Biochem. J.* **344**, 77–84 (1999).
50. Kraft, N. J. et al. Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* **29**, 592–599 (2015).
51. Abraham, Z. H., Lowe, D. J. & Smith, B. E. 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. *Biochem. J.* **295**, 587–593 (1993).
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. *J. Biochem.* **96**, 447–454 (1984).
53. Iwasaki, H. & Matsubara, T. A nitrite reductase from *Achromobacter cycloclastes*. *J. Biochem.* **71**, 645–652 (1972).
54. Liu, M. Y., Liu, M. C., Payne, W. J. & Legall, J. Properties and electron transfer specificity of copper proteins from the denitrifier '*Achromobacter cycloclastes*'. *J. Bacteriol.* **166**, 604–608 (1986).
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. *J. Biochem.* **89**, 453–461 (1981).
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).
57. Michalski, W. P. 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 Cell Physiol.* **19**, 1339–1351 (1978).
59. Denariaz, G., Payne, W. J. & LeGall, J. The denitrifying nitrite reductase of *Bacillus halodenitrificans*. *Biochim. Biophys. Acta Bioenerg.* **1056**, 225–232 (1991).
60. Glockner, A. B., Jüngst, A. & Zumft, W. G. Copper-containing nitrite reductase from *Pseudomonas aureofaciens* is functional in a mutationally cytochrome cd1-free background (NirS-) of *Pseudomonas stutzeri*. *Arch. Microbiol.* **160**, 18–26 (1993).
61. Timkovich, R., Dhesi, R., Martinkus, K. J., Robinson, M. K. & Rea, T. M. Isolation of *Paracoccus denitrificans* cytochrome cd1: comparative kinetics with other nitrite reductases. *Arch. Biochem. Biophys.* **215**, 47–58 (1982).
62. Gordon, E. H. 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. *J. Biol. Chem.* **278**, 11773–11781 (2003).
63. Besson, S., Carneiro, C., Moura, J. J., 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).
64. Sawhney, V. & Nicholas, D. J. Sulphide-linked nitrite reductase from *Thiobacillus denitrificans* with cytochrome oxidase activity: purification and properties. *J. Gen. Microbiol.* **106**, 119–128 (1978).

65. Stewart, V., Lu, Y. & Darwin, A. J. Periplasmic nitrate reductase (NapABC enzyme) supports anaerobic respiration by *Escherichia coli* K-12. *J. Bacteriol.* **184**, 1314–1323 (2002).
66. Van Alst, N. E., Sherrill, L. A., Iglewski, B. H. & Haidaris, C. G. Compensatory periplasmic nitrate reductase activity supports anaerobic growth of *Pseudomonas aeruginosa* PAO1 in the absence of membrane nitrate reductase. *Can. J. Microbiol.* **55**, 1133–1144 (2009).
67. Ikeda, E. et al. Physiological roles of two dissimilatory nitrate reductases in the deep-sea denitrifier *Pseudomonas* sp. strain MT-1. *Biosci. Biotechnol. Biochem.* **73**, 896–900 (2009).
68. Bell, L. C., Richardson, D. J. & Ferguson, S. J. Periplasmic and membrane-bound respiratory nitrate reductases in *Thiosphaera pantotropha*. The periplasmic enzyme catalyzes the first step in aerobic denitrification. *FEBS Lett.* **265**, 85–87 (1990).
69. Warnecke-Eberz, U. & Friedrich, B. Three nitrate reductase activities in *Alcaligenes eutrophus*. *Arch. Microbiol.* **159**, 405–409 (1993).
70. Rigoulet, M., Mourier, A., Galinier, A., Casteilla, L. & Devin, A. Electron competition process in respiratory chain: regulatory mechanisms and physiological functions. *Biochim. Biophys. Acta Bioenerg.* **1797**, 671–677 (2010).
71. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
72. Logares, R. et al. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.* **16**, 2659–2671 (2014).
73. Krueger, F. Trim Galore. *GitHub* https://github.com/FelixKrueger/TrimGalore (2023).
74. Bushnell, B., Rood, J. & Singer, E. BBmerge – accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
75. Cole, J. R. et al. The ribosomal database project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res.* **33**, D294–D296 (2005).
76. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
77. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
78. Wu, Y.-W., Simmons, B. A. & Singer, S. W. Maxbin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
79. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, 1–15 (2015).
80. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
81. 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).
82. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
83. Glöckner, F. O. et al. 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J. Biotechnol.* **261**, 169–176 (2017).
84. Yilmaz, P. et al. The SILVA and 'All-species Living Tree Project (LTP)' taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).
85. Waki, T., Murayama, K., Kawato, Y. & Ichikawa, K. Transient characteristics of Paracoccus denitrificans with changes between aerobic and anaerobic conditions. *J. Ferment. Technol.* **58**, 243–249 (1980).
86. Liu, P.-H., Zhan, G., Svoronos, S. A. & Koopman, B. Diauxic lag from changing electron acceptors in activated sludge treatment. *Water Res.* **32**, 3452–3460 (1998).
87. Liu, P.-H., Svoronos, S. A. & Koopman, B. Experimental and modeling study of diauxic lag of *Pseudomonas denitrificans* switching from oxic to anoxic conditions. *Biotechnol. Bioeng.* **60**, 649–655 (1998).
88. Gouw, M., Bozic, R., Koopman, B. & Svoronos, S. A. Effect of nitrate exposure history on the oxygen/nitrate diauxic growth of *Pseudomonas denitrificans*. *Water Res.* **35**, 2794–2798 (2001).
89. Lisbon, K., McKean, M., Shekar, S., Svoronos, S. A. & Koopman, B. Effect of do on oxic/anoxic diauxic lag of *Pseudomonas denitrificans*. *J. Environ. Eng.* **128**, 391–394 (2002).
90. Casasús, A. I., Hamilton, R. K., Svoronos, S. A. & Koopman, B. A simple model for diauxic growth of denitrifying bacteria. *Water Res.* **39**, 1914–1920 (2005).
91. *16S Metagenomic Sequencing Library Preparation*. Tech. Rep. (Illumina, 2013).
92. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
93. Callahan, B. J. et al. Dada2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
94. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with qiime 2's q2-feature-classifier plugin. *Microbiome* **6**, 1–17 (2018).
95. McLaren, M. R., Willis, A. D. & Callahan, B. J. Consistent and correctable bias in metagenomic sequencing experiments. *Elife* **8**, e46923 (2019).
96. Ku, H. H. et al. Notes on the use of propagation of error formulas. *J. Res. Natl Bur. Stand.* 70 (1966).
97. Taylor, J. R. *An Introduction to Error Analysis* 2nd edn (University Science Books, 1997).
98. Soil Survey Staff, USDA. Soil taxonomy: a basic system of soil classification for making and interpreting soil surveys. In *Agriculture Handbook* No. 436 (USDA, 1999).
99. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1–e1 (2013).
100. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
101. Douglas, G. M. et al. PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* **38**, 685–688 (2020).

Acknowledgements

We acknowledge L. M. de Jesús Astacio and M. Yousef for collecting soils, R. A. Oliveira for help with colony counting technique, M. Bahram for previously published global topsoil microbiome data and K. Husain for useful discussions. Soil sampling done in the CAF was, in part, a contribution from the 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 number 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 number 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.

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.K.L. and M.C.-W. performed the experiments. K.G. performed statistical analysis of the topsoil microbiome data and performed and analysed 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 consumer-resource model fits, predictions and simulations, adapting code written by K.G. K.K.L. sampled topsoils and performed the soil microcosm experiments. M.T. performed the analysis of sequencing data for the soil microcosm experiment in Supplementary Fig. 3. The topsoil microcosm experiment was, in part, a contribution from the LTAR network. LTAR is supported by the United States Department of Agriculture. K.C., K.G., Z.L., K.K.L., M.T. and S.K. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

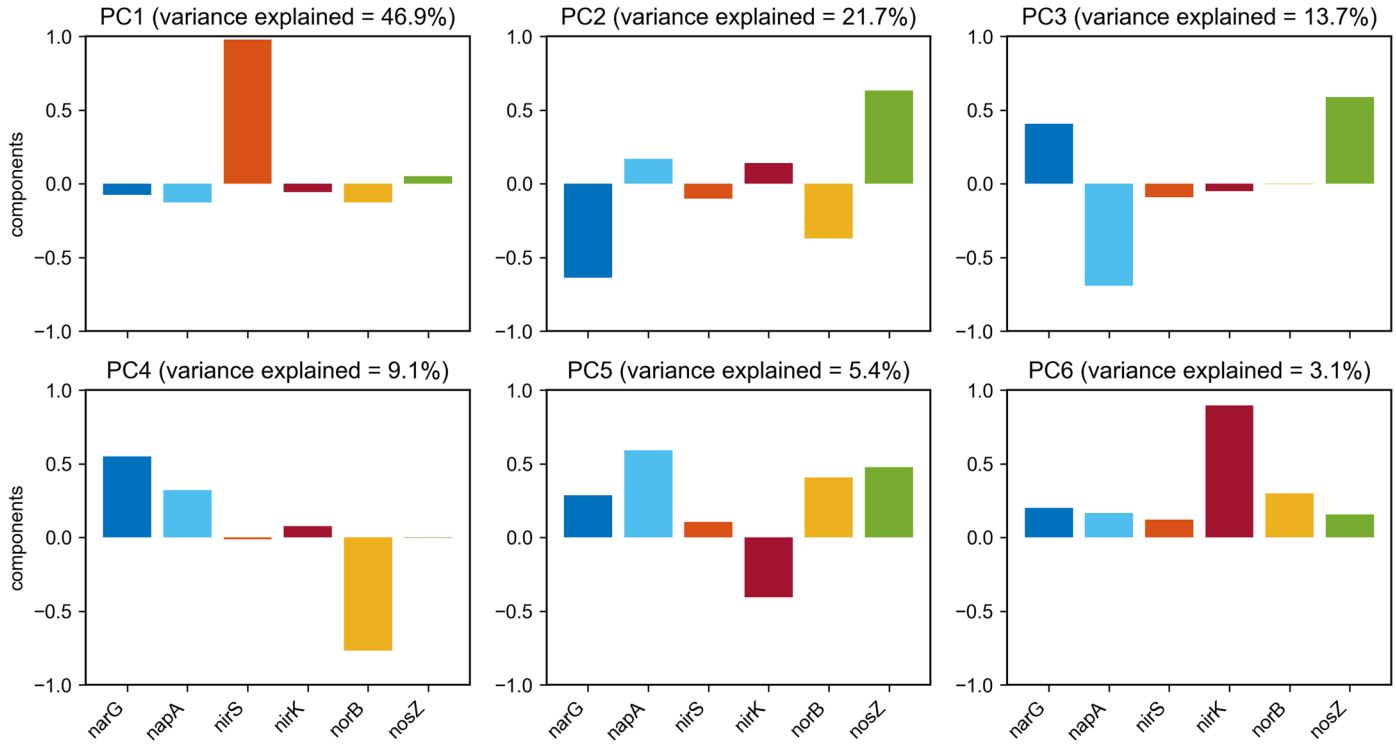
Extended data is available for this paper at https://doi.org/10.1038/s41564-024-01752-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-024-01752-4.

Correspondence and requests for materials should be addressed to Karna Gowda or Seppe Kuehn.

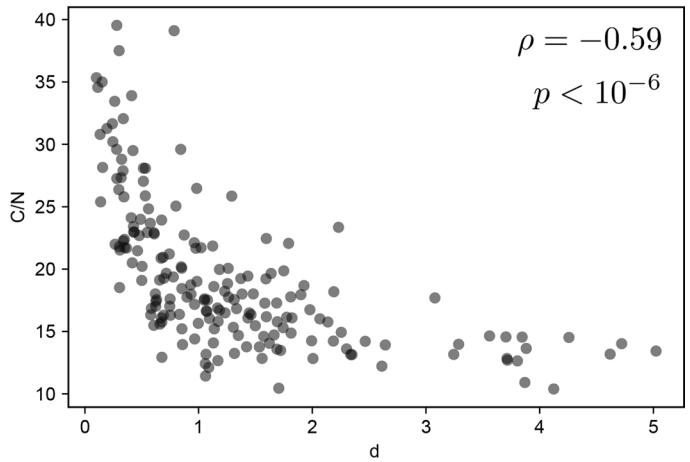
Peer review information *Nature Microbiology* thanks Nicholas Bouskill, Eoin Brodie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41564-024-01752-4



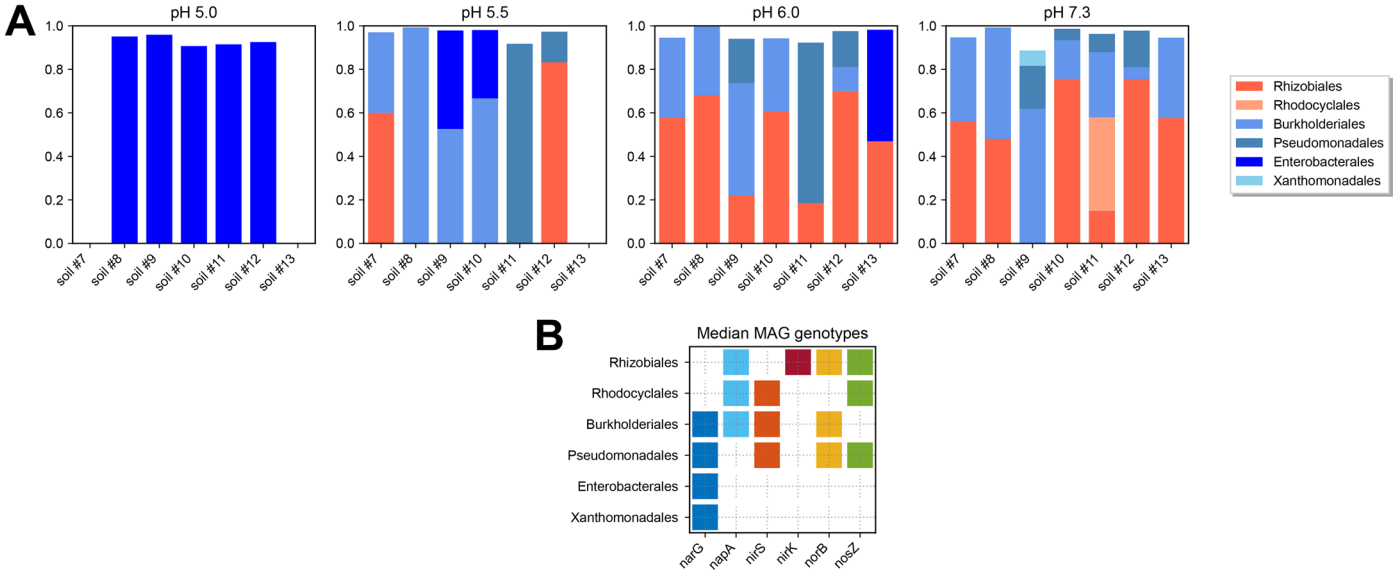
Extended Data Fig. 1 | Principal components resulting from the uiSVD decomposition of the global topsoil microbiome. Bars show the loadings of each gene in the six principal components (PCs) resulting from the uiSVD decomposition of denitrification reductase (*narG*, *napA*, *nirS*, *nirK*, *norB*, *nosZ*) relative abundances from the global topsoil microbiome.

https://doi.org/10.1038/s41564-024-01752-4



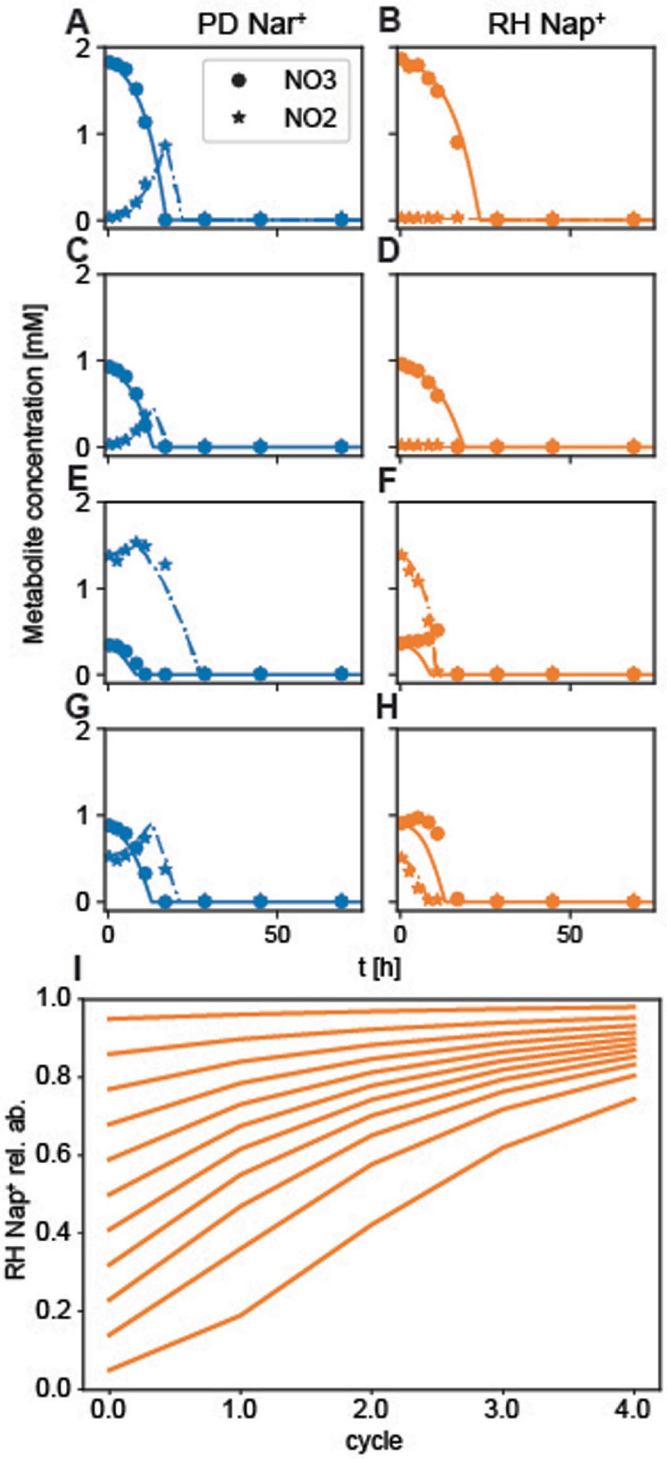
Extended Data Fig. 2 | C/N ratio correlates negatively with denitrification pathway magnitude. Unit-invariant singular value decomposition (uiSVD) was used to decompose denitrification reductase (*narG*, *napA*, *nirS*, *nirK*, *norB*, *nosZ*) relative abundances from the global topsoil microbiome into 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).

https://doi.org/10.1038/s41564-024-01752-4



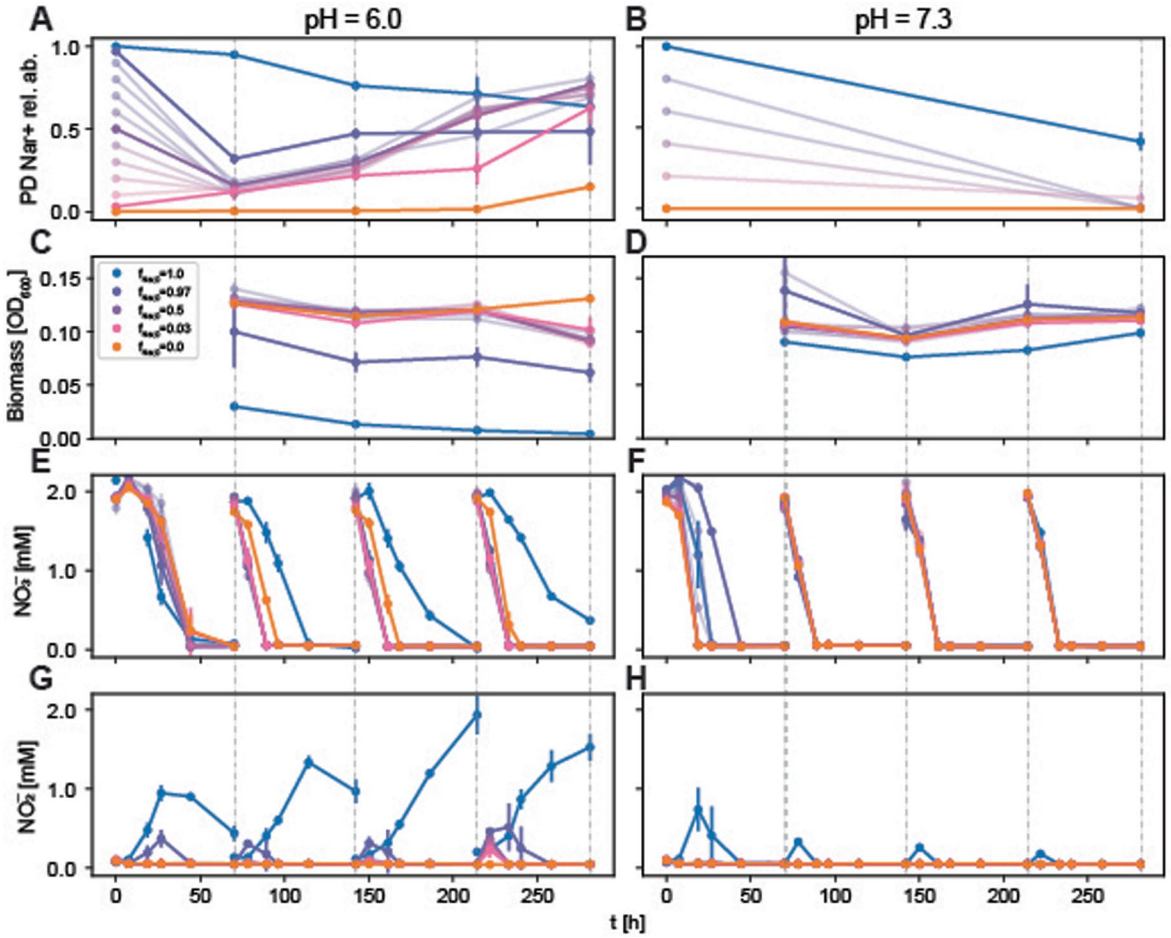
Extended Data Fig. 3 | Enrichment across a broader range of pH values. Additional enrichments were performed at pH 5.0, 5.5, 6.0, and 7.3 and the endpoint cultures were shotgun sequenced to infer taxonomic composition and genotypes. (**A**) Endpoint community compositions of the enrichments inferred 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.

https://doi.org/10.1038/s41564-024-01752-4



Extended Data Fig. 4 | pH 7.3 monoculture data are explained by individual fitness effects. (**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 ([NO^{-3}], [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).

https://doi.org/10.1038/s41564-024-01752-4



Extended Data Fig. 5 | Co-culture enrichment biomass and metabolite dynamics. Details of the PD Nar$^+$ and RH Nap$^+$ co-culture experiment shown in Fig. 5. (**A**) PD Nar$^+$ relative abundance dynamics at pH 6.0 are shown. $f_{0,Nar+} = 0, 0.03, 0.5, 0.97$ 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 invasible 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 $f_{0,Nar+} = 0$ 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 $f_{0,Nar+} = 0$ 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.

nature portfolio

Corresponding author(s): Seppe Kuehn and Karna Gowda

Last updated by author(s): 5/10/2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data for this study.
Data analysis	Paired-end reads from shotgun metagenomic sequencing were trimmed for quality using Trim-Galore (ver. 0.6.7). Taxonomic barcodes were extracted by merging overlapping paired end reads via BBMerge (ver. 38.22) and using the miTAGs extraction script (ver. 1). Taxonomic classification of miTAGs was performed on the RDP Classifier server (release 11, update 5). Metagenome assembled genomes were constructed using the metaWRAP pipeline (ver. 1.3): this pipeline includes assembly via metaSPAdes (ver. 3.15.4), binning using CONCOCT (ver.1.1.0), MaxBin2 (ver. 2.2.7), and metaBAT (ver. 2.15), and bin quality assessment using CheckM (ver. 1.0.11). MAGs were then annotated using the RAST server. Trimmomatic (ver. 0.39), SPAades (ver. 3.15.0), eggNOG-mapper (ver. 2.0.8), Minimap2, and the Kaiju web server (https://kaiju.binf.ku.dk/server) were used for phylogenetic classification of denitrification reductases. USEARCH (ver. 11.0), DADA2 (ver. 1.26), and PICRUSt2 (ver. 2.5.2) were used for sequence analysis of topsoil incubations. All other analyses and figures were generated in either Python (ver. 3.9.1) or MATLAB (ver. 2017b) using custom scripts. All scripts and necessary data to reproduce analyses in the manuscripts are available on our OSF repository: doi.org/10.17605/OSF.IO/N4J6F.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

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. (https://doi.org/10.1038/s41586-018-0386-6). The SILVA rRNA database (https://www.arb-silva.de) was used for taxonomic classification.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Our study combines the statistical analysis of previously published sequencing data of the global topsoil microbiome, the laboratory enrichment of microbial communities starting from natural soil samples, and the quantitative characterization of bacterial isolates from these enrichments.
Research sample	Our study comprised two primary research samples. (1) topsoils were sampled from areas around the Midwestern and Northwestern United States, with locations and dates for sampling listed in Tables S1 and S4. (2) Bacterial isolates were obtained from enrichment experiments inoculated with topsoil samples. Two isolates were studied in our manuscript, Rhizobium sp. RH Nap+ and Pseudomonas sp. PD Nar+.
Sampling strategy	No sample size calculations were performed for obtaining our research samples. Our goal for soil samples was to obtain a highly diverse collection of bacterial strains for initiating enrichment experiments. Our goal for bacterial isolation was to obtain representative taxa for enrichment experiments, which was assessed through comparison with sequencing data.
Data collection	Soil samples were collected using asceptic technique in order to minimize contact with non-endogenous microbes. Bacterial isolates were obtained by streaking cryopreserved samples of endpoint enrichment cultures on solid media, picking isolated colonies, and streaking to purity.
Timing and spatial scale	Timing and spatial scale are not applicable to our study design.
Data exclusions	No data were excluded from our study.
Reproducibility	Experimental designs for characterization of isolates included technical replicates to assess reproducibility, and the number of

Reproducibility	technical replicates is indicated where appropriate throughout the manuscript.
Randomization	Randomization is not applicable to the study design.
Blinding	Blinding is not applicable to the study design.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Soil samples were collected from natural areas in the Midwestern and Northwestern United States.
Location	Locations and dates for sampling locations are given in Tables S1 and S4.
Access & import/export	All sampling was performed on public land in compliance with local guidelines.
Disturbance	Disturbance was minimized by collecting small (< 100 g) soil samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals.
Reporting on sex	Reporting on sex is not applicable to the study.
Field-collected samples	Bacteria were isolated from laboratory enrichments which were inoculated with samples collected from forest preserves and manicured lawns using asceptic technique. Samples were stored separately for less than three months at 4 C before processing. Information about the collection dates and locations for each sample are given in Tables S1 and S4.
Ethics oversight	The protocols used in this study were approved by the University of Chicago Institutional Biosafety Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.