Genomic patterns in the global soil microbiome emerge from microbial interactions

Kyle Crocker,[1,2] Milena Chakraverti-Wuerthwein,[3] Zeqian Li,[1,2,4] Madhav Mani,[5,6,7] Karna Gowda,[1,2*] Seppe Kuehn[1,2*]

[1]Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637, USA
[2]Center for the Physics of Evolving Systems, The University of Chicago, Chicago, IL 60637, USA
[3]Biophysical Sciences Graduate Program, The University of Chicago, Chicago, IL 60637, USA
[4]Department of Physics, The University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[5]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA
[6]Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
[7]NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL 60208, USA

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

Microbial metabolism sustains life on Earth. Global sequencing surveys reveal ubiquitous patterns linking the taxa and genes responsible for metabolism with environmental conditions. One explanation for these patterns is environmental filtering: local conditions select for strains with particular traits. However, filtering assumes ecological interactions do not influence patterns. Here, we demonstrate a clear example of interactions driving global patterns in topsoil microbiomes.

We find a global trade-off between two nitrate reduction genes and pH: *nar* gene abundances increase while *nap* decreases with declining pH. Experiments contradict filtering by revealing strains possessing Nar thrive at low pH due to an interaction with Nap strains that mitigates nitrite toxicity. Thus, associations between environmental variables and gene abundances arise from environmentally modulated interactions driven by conserved metabolic traits.

Introduction

The metabolic activity of microbiomes in natural environments from soils to oceans drives 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 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. For example, bacterial metabolic capabilities vary with nutrient levels in marine systems (*2*), and gene content changes with soil pH (*3–5*) and host diet (*6–8*). Therefore, variation in the local environment modulates microbiome gene content and metabolic activity upon which the biosphere relies (*4, 9, 10*). Understanding the origins of environmentally-mediated variation in microbiome gene content is a necessity to understand how human activity impacts global nutrient cycles.

How do environmentally dependent patterns in gene content reproducibly emerge in microbial communities across the planet? One common hypothesis is environmental filtering: strains exhibiting traits that are adapted to their local environment are at a competitive advantage, and therefore increase in abundance. In some cases, associations between environmental variables and certain genotypes make intuitive sense with the environmental filtering hypothesis. For example, the relative abundance of acid-tolerant Acidobacteria in soil increases as the pH declines (*10–12*). Similarly, the prevalence of anaerobic nitrate respiration and anoxygenic photoautotrophy increase

with decreasing concentrations of oxygen in the global ocean microbiome (*2*). In these cases, we presume that specific metabolic strategies are enriched in niches where those traits facilitate a competitive advantage. However, we also know that microbial communities are complex systems with strong interactions between constituents (*13,14*), resulting in feedbacks, counter-intuitive inhibitory effects, and predation, all of which conspire to determine abundances (*15–19*). Thus environmental factors can also impact community diversity, gene content, and metabolic activity by modifying the nature of interactions.

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 consistent with the environmental filtering hypothesis is that interactions are weak, and changes in gene content reflect an adaptation of individual genotypes to local environmental conditions. A second possibility is that the patterns in gene content emerge from ecological interactions whose strength and specificity are modulated by local environmental variables. In this case, the emergence of reproducible patterns requires that the interactions exhibit regularity across different locations with similar environmental conditions. However, given the complexity of interactions in microbiomes, imagining how this regularity could manifest is a challenge.

Here, we present evidence that global patterns in the gene content of microbial communities emerge from environmentally-dependent interactions between members of the microbiome. We use bacterial denitrification as a model process to understand the mechanisms underlying global patterns in gene content. We quantify patterns in the abundances of genes associated with denitrification in the global topsoil microbiome (*4*), finding that soil pH strongly associates with variation in denitrification reductase gene content in soils. Using enrichments and quantitative phenotyping of isolates, we demonstrate the insufficiency of the environmental filtering hypothesis to explain these global patterns. Specifically, we show that strains possessing the denitrification gene Nar (Nar^+), encoding a cytoplasmic nitrate reductase, are globally enriched in low pH soils, and that strains possessing this gene dominate low pH enrichment cultures as well. Counterintuitively, Nar^+ strains

generically accumulate nitrite, which is toxic in acidic conditions, giving rise to reduced growth when cultured alone. We show that a strong interaction between Nar^+ strains and strains possessing the alternative nitrate reductase Nap reduces this toxicity, allowing Nar^+ to dominate at low pH. We argue that these interactions yield reproducible genomic patterns on a global scale as a result of the conserved physiological properties of the genotypes involved (*20*) and the interactions these physiological traits support. Our study provides a new unified way of understanding ubiquitous patterns in microbiomes as emerging from environmentally-modulated interactions between members of the community.

Results

Quantifying global patterns in topsoil microbiome gene content

We first sought to quantify variation in metabolic gene content on a global scale. To accomplish this, we focused on the soil microbiome because of the importance of these environments for global biogeochemistry, and the fact that global metagenomic surveys of topsoil microbiomes have recently been completed (*4*). Within the soil microbiome, we studied the process of denitrification, because it is ubiquitous, it arises from enzymes that are well-studied and reliably annotated (*21, 22*), and is performed by readily culturable taxa (*20, 23, 24*). As we demonstrate below, these properties enabled us to reliably identify genomic patterns in gene content from sequencing data and dissect interactions using wild isolates.

In soils, denitrifiers are present at substantial abundances ($\sim 1 \times 10^6$ cells/g, Ref. 25). These bacteria drive a cascade of reduction reactions ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$, Fig. 1A) to utilize organic carbon anaerobically via respiration. The cascade of reduction reactions is carried out by phylogenetically diverse bacteria (*26*), utilizing specific reductases to perform each reaction (Fig. 1A). The structure of the denitrification pathway varies between taxa, with some strains performing all four steps of the cascade, and others performing nearly arbitrary subsets of the four reactions (*24*). As such, denitrification is often a collective process in the wild.

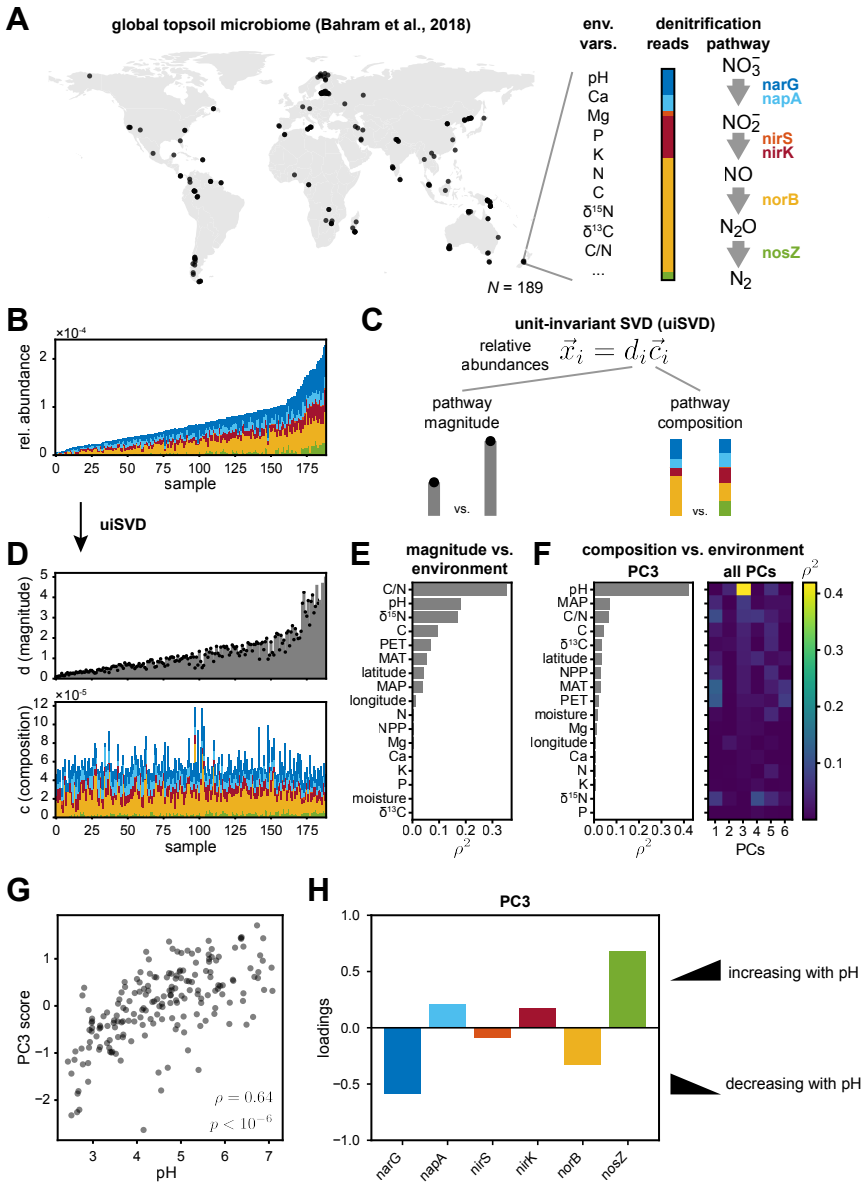

Figure 1: **pH is associated with covariation in denitrification pathway composition in the global topsoil microbiome.** **(A)** Topsoils sampled at 189 globally-distributed sites were chemically characterized (pH, Ca, Mg, ...), sequenced via shotgun metagenomics, and functionally annotated by Bahram et al. (*4*). **(B)** The relative abundance of denitrification reductases in each soil sample (relative to total gene content) are plotted in order of increasing total relative abundance. Reductase color legend indicated in panel A. **(C-D)** Unit-invariant singular value decomposition (uiSVD) (*27*) was used to decompose the data in panel B into contributions due to pathway magnitude (d_i) and pathway composition ($\vec{c}_i$). The results of this decomposition are plotted in panel **D**. **(E-F)** Pathway magnitude values and principal component (PC) scores for pathway composition obtained via uiSVD (Methods) are compared with 17 environmental variables, and squared Pearson correlation coefficients are shown. Pathway magnitude is most correlated with C/N ratio ($\rho = -0.59$, $p < 10^{-6}$ via randomization test; Fig. S1), while scores of PC3 are most correlated with pH ($\rho = 0.64$, $p < 10^{-6}$; **G**). **(H)** Loadings of PC3 are shown, where positive values indicate reductase content that increases with pH, and vice versa.

To study denitrification gene content in soils we utilized the Bahram et al. (*4*) survey of the topsoil microbiomes. The dataset comprises 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 relate environmental factors to the abundances of genes encoding denitrification enzymes.

First, we quantified variation in denitrification gene content across soils in the global topsoil microbiome. To do this, we computed the fraction of reads in each sample that map to any of the six main reductases in the denitrification pathway (*narG*, *napA*, *nirS*, *nirK*, *norB*, and *nosZ*; Fig. 1A). The result is a table that contains the fraction of all reads mapping to any of the reductases in the denitrification pathway (y-axis, Fig. 1B). We refer to this total fraction as the "magnitude" of the pathway (Fig. 1C). Our analysis also quantifies the pathway composition, defined as the relative contribution of each reductase to the total (colors, Fig. 1B). Denitrification pathway magnitude varies by about 50-fold across soil samples, while pathway composition also varies from sample to sample, but this variation is challenging to visually distinguish from variation in pathway magnitude (Fig. 1B).

To disentangle pathway magnitude from composition, we developed a new approach based on unit-invariant singular value decomposition (uiSVD, (*27*); Methods). uiSVD decomposes 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. 1C). The result of applying this method to the denitrification reductase relative abundance matrix is shown in Fig. 1D. Next, we sought to determine how environmental variables are associated with the magnitude and composition of the denitrification pathway in soils.

pH correlates with pathway composition on a global scale

To connect environmental variation to patterns in pathway magnitude, we examined correlations between environmental variables and magnitude. We considered squared Pearson correlations (ρ^2) between pathway magnitudes (d_i) and each environmental variable in turn. We found that the magnitude of the pathway was most strongly associated with the C/N ratio ($\rho^2 = 0.35$, $p < 10^{-6}$; Fig. 1E). The correlation is negative ($\rho = -0.59$; Fig. S1), consistent with previous work showing that denitrification is enhanced at low C/N ratios (*28*).

To understand how environmental variation impacts the prevalence of the different denitrification enzymes, we leveraged the fact that uiSVD yields the principal components (PCs) of the composition matrix (Methods), which represent covarying groups of genes. By looking at correlations between PC scores (projections of composition onto the PCs) to the environmental variables, we found that a single component of variation, PC3, was associated with pH much more strongly than any other combination of components and environmental variables ($\rho^2 = 0.42$, $p =< 10^{-6}$; Fig. 1F,G). Examining the loadings of each reductase in PC3 (Fig. 1H) reveals how denitrification reductase genes vary with pH on a global scale: some reductase genes are enriched as pH decreases (*narG*, *nirS*, *norB*), while others are depleted as pH decreases (*napA*, *nirK*, *nosZ*). While pH is widely appreciated as a strong contributor to community diversity in soils (*9*), and pH is known to be important for denitrification (*29, 30*), our analysis provides a quantitative view of how pH impacts which genes are involved in denitrification on a global scale. Our approach provides a principled, generalizable route to discerning patterns in large-scale microbiome surveys.

What causes some gene abundances to rise and others to fall with changes in pH? A strict environmental filtering hypothesis would suggest that strains possessing specific reductases are favored in soils of different pH. Under this assumption, Fig. 1F suggests the hypothesis that strains possessing the nitrate reductase Nar are favored at low pH, while those possessing the nitrate reductase Nap are favored at neutral pH. However, to our knowledge there is no direct evidence for this hypothesis, nor is this hypothesis easily tested from sequencing data alone. Alternatively, it

is also possible that pH is not causally associated with variation in reductase content, but rather related to unmeasured causally-associated variables. Therefore, we sought to test the dependence of denitrification genotypes on pH experimentally using an enrichment culture approach.

Enrichment cultures recapitulate global patterns in denitrification gene content

We reasoned that if pH is causally related to genotype abundances as shown in Fig. 1F, we should be able to recapitulate these patterns in the lab. To experimentally test the associations between pH and denitrification genotypes, we performed enrichment cultures under different pH conditions (Fig. 2A). We extracted diverse microbial communities from six soil samples (Table S1), and inoculated them into chemically defined media buffered at pH 6.0 and 7.3. After a $72\,\mathrm{h}$ growth period in anaerobic conditions, we passaged these cultures into fresh medium ($8-$fold dilutions). We repeated this passaging serially for 12 total $72\,\mathrm{h}$ cycles, supplying $1.75\,\mathrm{mM}$ nitrate (7 parts $2\,\mathrm{mM}$ fresh media to 1 part spent media) at the beginning of each cycle. We chose these conditions because pH in soils is often strongly buffered (*31–33*) and mM concentrations of nitrate are often present (*34, 35*). The resulting enrichment cultures were then cryopreserved and shotgun sequenced to assay taxonomic and functional metagenomic composition. The details of all culturing and sequencing procedures can be found in Methods.

First, to assay taxonomic composition of the endpoint enrichment cultures, we extracted 16S rRNA miTAGs (*36*) from shotgun metagenome sequencing data and annotated them (*39*) (Methods). This analysis revealed that endpoint enrichments were dominated by taxa from the orders Pseudomonadales and Rhizobiales, with the former comprising the majority of all six 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 at 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 in any enrichment at an abundance greater than 1%.

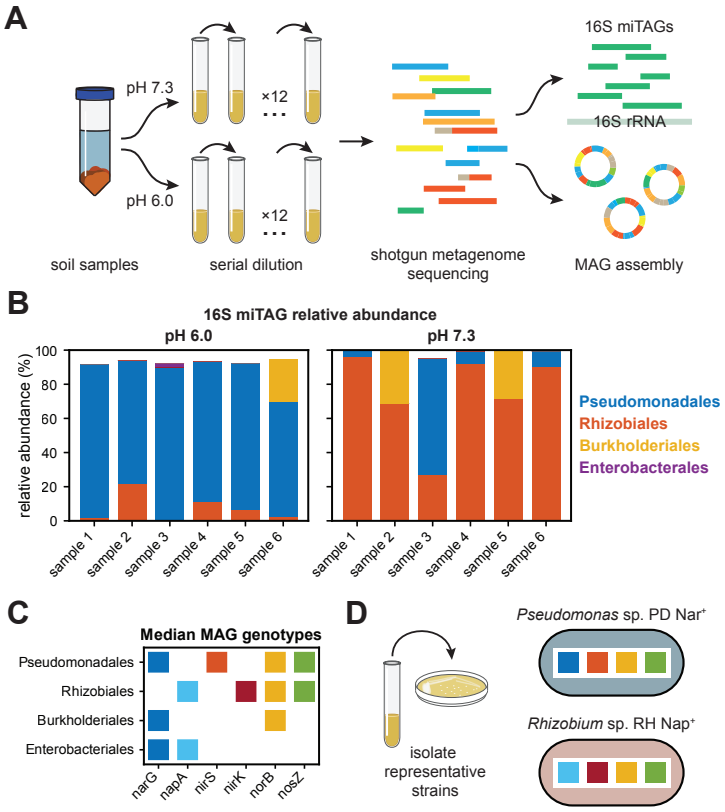

Figure 2: **Enrichment cultures recapitulate global patterns in denitrification gene content.** **(A)** Denitrifying communities were enriched from soil samples in two pH conditions (6.0 and 7.3), and then sequenced via shotgun metagenomics to measure taxonomic composition and genotypes. Six soil samples were mechanically homogenized and used to inoculate a serial dilution experiment in denitrifying (anaerobic) conditions. After 72 h growth cycles, cultures were repeatedly passaged (12×) into a defined medium containing 2 mM nitrate via a 1/8 dilution factor (Methods). **(B)** Endpoint community compositions are shown at the level of taxonomic order for enrichments in pH 6.0 and 7.3. Compositions were determined by extracting 16S rRNA fragments (miTAGs) from shotgun metagenome data (*36*), and then taxonomically annotating miTAGs via RDP (*37*). Taxa present at a relative abundance less than 1% are omitted. **(C)** Median denitrification reductase genotypes are shown for metagenome-assembled genomes (MAGs) corresponding to the four most abundant taxa present in the composition data in panel B. MAGs were functionally annotated using RAST (*38*), and the median genotype was computed over MAGs obtained in different samples. **(D)** Strains were isolated from cryopreserved samples from the enrichment endpoint (Methods). The strain *Pseudomonas* sp. PD Nar$^+$ represents the dominant taxa present across samples at pH 6.0, and *Rhizobiales* sp. RH Nap$^+$ represents the dominant taxa across samples at pH 7.3. See also Figs. S2, S3.

To determine the genotypes of taxa in the enrichment experiments, we constructed metagenome-assembled genomes (MAGs) (*40*) (Methods). Annotating these MAGs revealed that taxa from different enrichments had largely similar denitrification genotypes (Fig. 2C); in particular, Pseudomonadales MAGs typically encoded the reductases NarG and NirS, while Rhizobiales MAGs encoded the NapA and NirK reductases. A second enrichment experiment, performed with additional soil samples enriched at a broader range of pH values, also recapitulated the same basic pattern in genotype abundances, with taxa possessing NarG heavily enriched at pH 5.0 and 5.5 (Fig. S3; Methods).

Strikingly, these genotypes recapitulate the patterns in denitrification gene content observed in the topsoil microbiome (Fig. 1F); namely, our enrichment resulted in a greater prevalence of *narG* and *nirS* at pH 6.0, and alternatively a greater prevalence of *napA* and *nirK* at pH 7.3. This agreement between the enrichment and our analysis of the global topsoil microbiome suggests that (1) pH is likely causally related to the patterns of gene covariation described in Fig. 1F, and (2) the mechanisms shaping these patterns in nature are also at play *in vitro*. Altogether, this points to the possibility that we can uncover the mechanistic basis for global patterns of gene-environment covariation by interrogating our enrichment experiments of community dynamics and assembly.

Environmental filtering does not explain outcome of acidic enrichments

To uncover the mechanisms that give rise to the observed pattern between denitrification gene content and soil pH (Fig. 1), we isolated taxa from the endpoint enrichment experiment from the genera *Pseudomonas* and *Rhizobium* (Fig. 2D, Methods). These strains correspond to the Nar^+ and Nap^+ genotypes, respectively (Fig. 2D). We denote the taxonomy and genotype of these isolates as PD Nar^+ and RH Nap^+.

The environmental filtering hypothesis would suggest that the Nar^+ genotype should thrive in acidic conditions and the Nap^+ genotype in neutral conditions. To test this hypothesis, we performed a serial dilution experiment with strains grown in isolation (Methods). We measured the

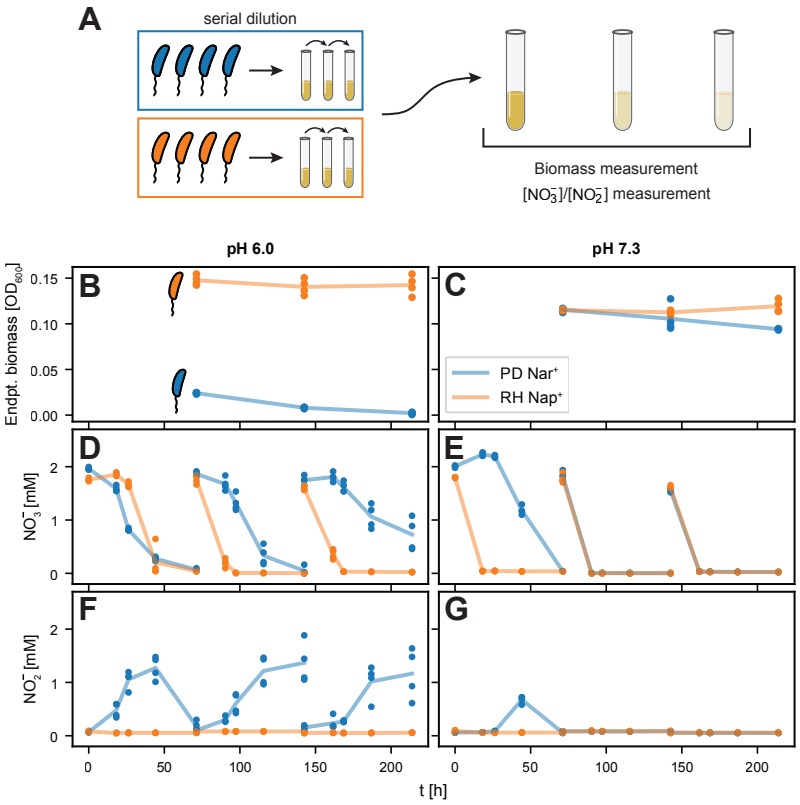

Figure 3: **Environmental filtering does not explain outcome of acidic enrichments.** **(A)** Schematic of the experimental design. PD Nar$^+$ (blue) and RH Nap$^+$ (orange) were passaged repeatedly in monoculture under denitrifying (anaerobic) conditions. After each $72\,\mathrm{h}$ growth cycle, cultures were passaged ($3\times$) into a defined medium containing $2\,\mathrm{mM}$ nitrate using a $1/8$ dilution factor. Biomass is measured via $600\,\mathrm{nm}$ absorbance at the end of each cycle, and nitrate and nitrite concentrations are measured throughout each cycle using a Griess assay (Methods). (**B** and **C**) Endpoint biomass is shown for each strain at both pH 6.0 (panel **B**) and 7.3 (panel **C**). Lines connect the average across replicates. RH Nap$^+$ (orange) produces more biomass at both pH levels, while PD Nar$^+$ (blue) appears to decay in abundance at pH 6.0 (despite PD Nar$^+$ community enrichment in this condition, Fig. 2B). (**D-G**) Nitrate and nitrite concentration dynamics are shown at six timepoints throughout each cycle in each condition. PD Nar$^+$ nitrate and nitrite reduction rates slow at pH 6.0 as the growth-dilution cycles progress, consistent with its reduction in biomass (Panels **D** & **F**, blue). PD Nar$^+$ (blue) accumulates nitrite at each pH condition, whereas RH Nap$^+$ (orange) does not (Panels **F** & **G**).

temporal dynamics of nitrate, nitrite, and total biomass at the end of each cycle in this experiment (Fig. 3).

In line with the environmental filtering hypothesis, the strain RH Nap$^+$ (representative of taxa most abundant in pH 7.3 enrichments) appeared to perform better than PD Nar$^+$ at pH 7.3. RH Nap$^+$ produced more biomass and consumed nitrate and nitrite more quickly in the first cycle (Fig. 3C, E, G). Additionally, simulated competition between these strains confirms that RH Nap$^+$ is expected to be enriched at pH 7.3 based on monoculture phenotype (Supplementary Information).

In contrast, results at pH 6.0 were not consistent with the environmental filtering hypothesis. PD Nar$^+$ produced only a small amount of biomass, with levels decreasing with each cycle, suggestive of cell death (Fig. 3B). At the same time, this strain reduced nitrate more slowly than RH Nap$^+$ in cycles 2–4 and reduced nitrite more slowly in all cycles (Fig. 3D, F). Crucially, PD Nar$^+$ transiently accumulated nitrite (blue curves, Fig. 3F), 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 is inconsistent with the environmental filtering hypothesis, suggesting that the dominance of Nar$^+$ genotypes in acidic enrichment experiments emerged from community-level processes, which are as yet unknown.

Nitrite toxicity impacts denitrification activity of isolates at low pH

As a route to understanding how community-level processes account for the discrepancy between soil enrichments (Fig. 2) and monocultures (Fig. 3), we next sought to uncover the mechanisms underlying the poor performance of PD Nar$^+$ monocultures at pH 6.0. Given the accumulation of nitrite in this experiment (Fig. 3F), as well as literature indicating the toxicity of nitrite under acidic conditions (*41–43*), we hypothesized that nitrite accumulation inhibited the growth of PD Nar$^+$ at pH 6.0.

We tested whether nitrite inhibited the metabolic activity of either isolate at pH 6.0. To do this,

we initiated monocultures of each strain at pH 6.0, supplying $1.75\,\mathrm{mM}$ nitrate and varying concentrations of nitrite. We then measured nitrate and nitrite dynamics over $72\,\mathrm{h}$ of growth (Fig. 4A). We found that the extent of inhibition increased with initial nitrite concentration for both strains (Fig. 4A). Interestingly, nitrite appeared to inhibit these two strains in qualitatively different ways. The nitrate reduction rate for PD Nar$^+$ slowed in a manner that was continuous with initial nitrite concentration (Fig. 4A), while nitrate reduction rate for RH Nap$^+$ showed a threshold-like dependence on initial nitrite concentration ($[\mathrm{NO}_2^-] \sim 0.4\text{-}0.5\,\mathrm{mM}$). Altogether, these results affirm that the accumulation of nitrite under acidic conditions (blue curves, Fig. 3F) should particularly impact PD Nar$^+$, while in contrast, RH Nap$^+$ does not accumulate nitrite appreciably (orange curves, Fig. 3F) and therefore should avoid inhibition.

Next, we sought to determine whether nitrite toxicity caused cell death. To do this, we initialized monocultures at pH 6.0, with no nitrate, and 0, 0.875, or $1.75\,\mathrm{mM}$ nitrite (Methods). In these conditions, we observed no cell growth, allowing us to quantify mortality via plating (Methods). The density of colony forming units (CFU/mL) as a function of time is shown for both strains in Fig. 4B. The dark purple line ($0\,\mathrm{mM}$ nitrite) is a control for stationary-phase mortality in the absence of nitrite. Cell densities below these values indicate death due to nitrite toxicity. We observed the mortality of PD Nar$^+$ increased sharply with increasing nitrite concentration at pH 6.0, while the mortality rate of RH Nap$^+$ changes little with nitrite concentration. This result strongly suggests that the accumulation of nitrite by PD Nar$^+$ in monoculture (Fig. 3F) drives cell death.

Finally, we compared the growth of PD Nar$^+$ in pH 6.0 monocultures supplied with different initial concentrations of nitrate. Because nitrite is the direct product of nitrate reduction, lower initial concentrations of nitrate result in less nitrite accumulation, and in principle less toxicity under acidic conditions. As expected, experiments initialized with either $1.75\,\mathrm{mM}$ (high) or $0.875\,\mathrm{mM}$ (low) nitrate indeed resulted in different degrees of nitrite accumulation (light/dark blue data points, Fig. 4C). Measuring endpoint biomass strongly indicated that the degree of toxicity was greater in the high nitrate condition. PD Nar$^+$ produced substantially less biomass in the high nitrate con-

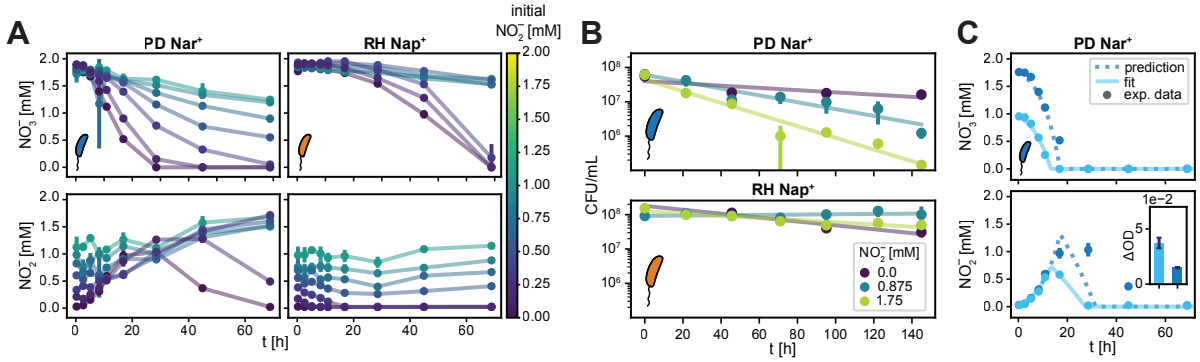

Figure 4: **Nitrite toxicity impacts denitrification activity of isolates at low pH.** (**A**) PD Nar$^+$ and RH Nap$^+$ were inoculated in anaerobic monoculture at pH 6.0 with $1.75\,\mathrm{mM}$ nitrate and varying nitrite levels (colorbar). Nitrate and nitrite concentration dynamics, shown for each strain, were measured via Griess assay (Methods). As the initial nitrite concentrations ($[\mathrm{NO}_2^-]_0$) rose, the growth of both strains was inhibited. PD Nar$^+$ was unable to fully reduce nitrate when $[\mathrm{NO}_2^-]_0 > 0.35\,\mathrm{mM}$ and was unable to fully reduce nitrite when $[\mathrm{NO}_2^-]_0 > 0\,\mathrm{mM}$. Similarly, RH Nap$^+$ was unable to fully reduce either nitrate or nitrite when $[\mathrm{NO}_2^-]_0 > 0.35\,\mathrm{mM}$. (**B**) PD Nar$^+$ and RH Nap$^+$ were again grown in anaerobic monoculture at pH 6.0 with varying nitrite levels, but nitrate was not supplied to prevent growth (Methods). Density of colony forming units was measured via plating. Lines are log-linear fits, and the (negative) slope of the line indicates mortality rate (Methods). For PD Nar$^+$, inferred death rates were 0.008 ± 0.004, 0.023 ± 0.004, and $0.039 \pm 0.004\ \mathrm{h}^{-1}$ for $[\mathrm{NO}_2^-]_0 = 0$, 0.875, and $1.75\,\mathrm{mM}$, respectively. For RH Nap$^+$, inferred death rates were 0.013 ± 0.004, -0.001 ± 0.001, and $0.007 \pm 0.001\ \mathrm{h}^{-1}$ for $[\mathrm{NO}_2^-]_0 = 0$, 0.875, and $1.75\,\mathrm{mM}$, respectively. (**C**) Nitrate (top) and nitrite (bottom) metabolite dynamics for monocultures of PD Nar$^+$, supplied with initial nitrate concentrations of $0.875\,\mathrm{mM}$ (light blue) and $1.75\,\mathrm{mM}$ (dark blue). A consumer resource model was fit to the low initial nitrate data (solid light blue lines) and used to predict high initial nitrate data (dashed dark blue lines). The prediction failed at intermediate time points (bottom panel), indicating that the accumulation of nitrite slowed the metabolic rate. Inset in bottom panel shows final optical density (OD). Endpoint biomass levels were greater in the $0.875\,\mathrm{mM}$ nitrate condition than in the $1.75\,\mathrm{mM}$ condition, suggestive of mortality induced by nitrite accumulation.

dition (inset, Fig. 4C), amounting to one-fifth of the biomass yield per unit substrate consumed compared with the low nitrate condition.

To further quantify the effect of nitrite toxicity in these experiments, we fit a consumer-resource model (CRM) to the metabolite dynamics and final biomass in the low nitrate condition. The CRM quantifies metabolite dynamics and growth using a Monod equation formalism and has been shown to describe denitrification dynamics reliably (*20*). We fit CRM parameters to the low nitrate condition and used these parameters to predict the high nitrate condition (Methods). If PD Nar$^+$ experienced the same degree of toxicity in both conditions, then the fits of the CRM to the low nitrate condition should accurately predict the high nitrate condition. In contrast, a greater degree of toxicity in the high nitrate condition would be reflected in poor model predictions. Indeed, we observed that predictions of the high nitrate condition were poor (dark blue, Fig. 4C); nitrate was consumed more slowly than predicted, indicating a greater degree of toxicity in the high nitrate condition.

Overall, these three experiments demonstrated that nitrite accumulation has a significant deleterious effect on the growth and metabolic activity of PD Nar$^+$.

Community alleviates nitrite toxicity under acidic conditions

PD Nar$^+$ failed to grow in monoculture at pH 6.0 (Fig. 3) due to toxicity from accumulating nitrite (Fig. 4), but still managed to dominate the community during enrichment in these conditions (Fig. 2). We hypothesized that since RH Nap$^+$ avoids nitrite accumulation in monoculture (dark purple traces, Fig. 4A; orange traces, Fig. 3F), its presence in a co-culture with PD Nar$^+$ could serve to reduce nitrite accumulation, alleviating toxicity and allowing PD Nar$^+$ to dominate.

To test the hypothesis that the presence of RH Nap$^+$ in co-culture alleviates toxicity, we grew a 1:1 mixture (by biomass) of PD Nar$^+$ and RH Nap$^+$ at pH 6.0, supplying $1.75\,\mathrm{mM}$ nitrate (Methods). We found this co-culture accumulated substantially less nitrite relative to the monoculture of PD Nar$^+$, producing significantly more biomass (inset, Fig. 5B). Remarkably, nitrite levels in

co-culture reached $\sim 0.5\,\mathrm{mM}$, which is at or below the toxicity threshold identified in earlier experiments (Fig. 4A).

We conclude that the combination of PD Nar$^+$ and RH Nap$^+$ experienced substantially less toxicity due to nitrite than PD Nar$^+$ by itself. We note that this community effect appears to derive from the apparent specializations of the two isolates on the resources nitrate and nitrite: PD Nar$^+$ exhibited fast nitrate reduction and slow nitrite reduction, while RH Nap$^+$ showed the opposite behavior (dark purple traces, Fig. 4A). While the nitrate specialization of PD Nar$^+$ causes toxicity because it results in nitrite accumulation, the nitrite specialization of RH Nap$^+$ allows it to remediate this toxicity.

We next turned to demonstrating that diminished nitrite toxicity in co-cultures of PD Nar$^+$ and RH Nap$^+$ explains the long-term dominance of PD Nar$^+$ across multiple cycles of growth and dilution (Fig. 2B). To do this, we co-cultured these isolates over four cycles of serial dilution and measured relative abundances via 16S amplicon sequencing (Fig. 5D; Methods). In order to assess whether density-dependent effects determine the outcome of these experiments, we varied initial relative abundances of PD Nar$^+$ and RH Nap$^+$ over a wide range, holding total biomass constant.

Indeed, we found that after four cycles at pH 6.0, co-cultures across a range of initial relative abundances 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 (*44*). 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 was increased with the relative abundance of PD Nar$^+$, with low abundances of PD Nar$^+$ showing very little nitrite accumulation (orange/pink traces, Fig. S7G), and vice versa (blue/purple traces, Fig. S7G). This, along with the earlier results showing that RH Nap$^+$ alleviates nitrite toxicity, supports the interpretation that PD Nar$^+$ requires the presence of RH Nap$^+$ at pH 6.0 to dominate.

Similar co-cultures performed at pH 7.3 showed RH Nap$^+$ driving PD Nar$^+$ to extinction after

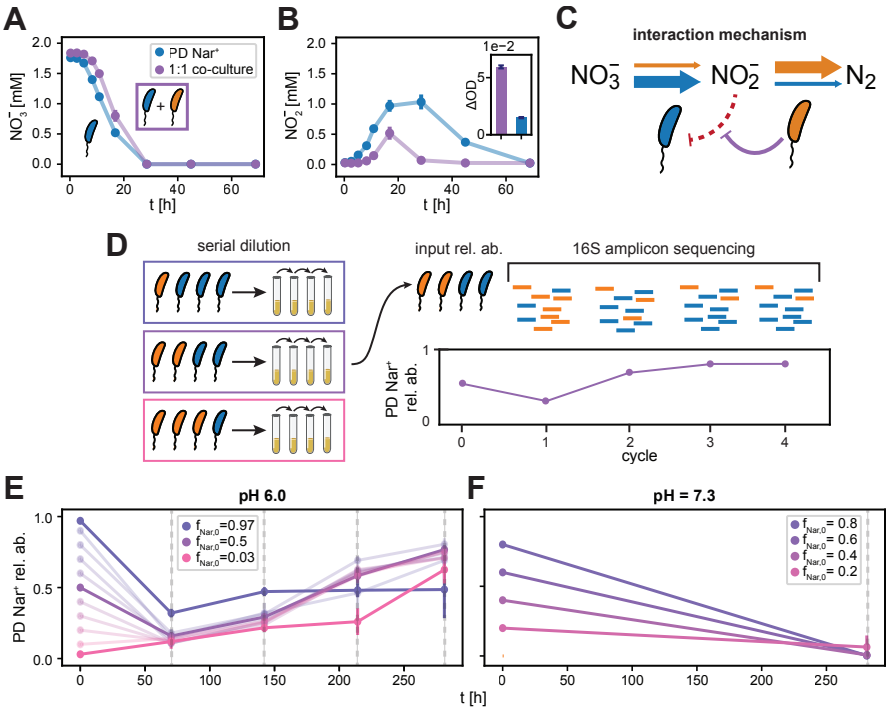

Figure 5: **Community alleviates nitrite toxicity under acidic conditions.** (**A-B**) Monoculture PD Nar$^+$ growth metabolite dynamics (shown previously in Fig. 4C) compared with the those of a 1:1 PD Nar$^+$ and RH Nap$^+$ co-culture at pH 6.0. Panel **A** shows nitrate concentration and **B** shows nitrite concentration. Initial biomass ($OD_{600} = 0.01$) and nitrate ($1.75\,\mathrm{mM}$) was the same in each condition, but the co-culture (purple) exhibited significantly less nitrite accumulation despite only slightly decreased nitrate reduction rate, as well as significantly higher biomass production (purple bar, inset, panel B). (**C**) Schematic of proposed interaction between PD Nar$^+$ and RH Nap$^+$. PD Nar$^+$ (blue cell) reduces nitrate quickly and accumulates nitrite (thick blue arrow). In monoculture, this nitrite accumulation is self-inhibitory (dashed red line). In co-culture, however, RH Nap$^+$ (orange cell) quickly reduces nitrite (thick orange arrow), alleviating the inhibitory effects of nitrite on PD Nar$^+$ (purple line). (**D**) Schematic illustrating the multi-cycle co-culture experiment shown in panels **E** and **F**. Mixtures of PD Nar$^+$ and RH Nap$^+$ were prepared across a range of ratios spanning 0.03:0.97 to 0.97:0.03 (distinguished by color), with total biomass held constant. These mixtures were then transferred to fresh media buffered at pH 6.0 or pH 7.3, with $2\,\mathrm{mM}$ nitrate supplied. Cultures were grown under anaerobic conditions for $72\,\mathrm{h}$ 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. (**E**) PD Nar$^+$ relative abundance dynamics at pH 6.0 are shown. $f_{Nar^+,0} = 0.03, 0.5,$ and 0.97 are highlighted by darker lines. After a transient decrease in the $f_{Nar^+,0} > 0.1$ conditions, co-cultures approach a value of $f_{Nar^+} \approx 0.60$–0.75. (**F**) PD Nar$^+$ relative abundance dynamics at pH 7.3 are shown. The relative abundance of PD Nar$^+$ approached zero after four cycles in all conditions, indicating that it was competitively excluded by RH Nap$^+$ in this pH condition.

four cycles (Fig. 5F), corroborating our expectations from earlier experiments and analyses that RH Nap$^+$ is simply a better competitor for resources at this pH condition (Fig. 3C, E, G; Fig. S6).

Nar$^+$ and Nap$^+$ phenotypes are conserved across diverse taxa

Using isolates from our soil enrichment experiments (Fig. 2D), we demonstrated (Figs. 3-5) that the dominance of a Nar$^+$ genotype (PD Nar$^+$) under acidic conditions required the presence of a Nap$^+$ genotype (RH Nap$^+$). Underlying the interactions between these isolates was the fact that the former is an apparent nitrate specialist, capable of quickly consuming the primary limiting resource supplied to the community, while the later is a nitrite specialist, which mitigates the toxic effects of accumulated nitrite. Generalizing this observation, we next hypothesized that patterns of genotypes in the global topsoil microbiome (Fig. 1) arose due to similar interactions driven by the same patterns of specialization: namely, that Nar$^+$ and Nap$^+$ display nitrate and nitrite specialist phenotypes, respectively, meaning that the former genotypes tend to accumulate nitrite while the latter alleviate its toxicity.

First, we asked whether or not it is typical for denitrifiers to possess either Nar or Nap independently, rather than possessing both. An analysis of over 2000 denitrifier genomes in the KEGG reference genome database confirmed that this is indeed the case (Supplementary Information), showing a negative correlation between the presence/absence of these two genes (Fig. S8).

Second, we asked whether the traits and resource specializations of these genotypes are conserved across taxa. Recent work showed that the rates of nitrate and nitrite utilization by diverse denitrifying bacteria are strongly correlated with the presence and absence of denitrification genes they possess (*20*). As a result, we expect that Nar$^+$ and Nap$^+$ genotypes generically exhibit the phenotypes shown in Fig. 3.

To answer this question, we leveraged the library of phylogenetically diverse Nar$^+$ and Nap$^+$ strains isolated in our previous study (*20*). We identified 7 representative strains possessing one or the other nitrate reductase (3 Nar$^+$, 4 Nap$^+$) spanning α, β and $\gamma-$proteobacteria. For each strain,

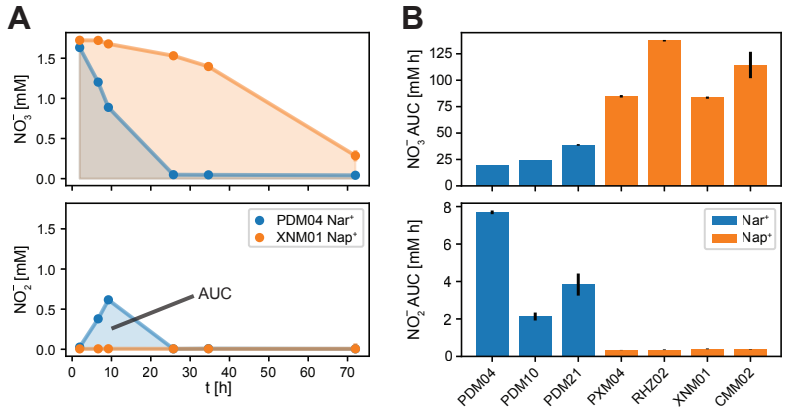

Figure 6: **Nar$^+$ and Nap$^+$ phenotypes are conserved across diverse taxa.** Additional Nar$^+$ and Nap$^+$ soil isolates were grown in anaerobic conditions at pH 6.0, and their denitrification phenotypes were measured. (**A**) Denitrification phenotypes were summarized by area under the nitrate (top) and nitrite (bottom) curves (AUC). Metabolite dynamics for PDM04 Nar$^+$ (blue) and XNM01 Nap$^+$ (orange) monocultures are shown, with area under the curve (AUC) for each metabolite illustrated by the colored shaded regions. Smaller AUC indicates faster metabolite reduction. (**B**) AUC for nitrate (top) and nitrite (bottom) for three Nar$^+$ and four Nap$^+$ strains isolated in Ref. (*20*). Nar$^+$ strains exhibit faster nitrate reduction (blue bars, top panel) and slower nitrite reduction (blue bars, bottom panel), and Nap$^+$ strains the opposite.

we measured nitrate and nitrite utilization phenotypes at pH 6.0 supplied with 1.75 mM nitrate, using the area under the curve (AUC) of nitrate and nitrite dynamics to summarize traits (Fig. 6A). Values of nitrate AUC are inversely proportional to the rate of nitrate consumption (top panel, Fig. 6A), while values of nitrite AUC are directly proportional to nitrite accumulation (bottom panel, Fig. 6A).

As with PD Nar$^+$ and RH Nap$^+$, we found that Nar$^+$ genotypes all exhibited fast nitrate reduction and the transient accumulation of nitrite (blue bars, Fig. 6B). In contrast, Nap$^+$ strains all exhibited slow nitrate reduction and no nitrite accumulation (orange bars, Fig. 6B). We conclude that Nar$^+$ genotypes generically accumulate nitrite via fast nitrate reduction, and that Nap$^+$ strains generically avoid this accumulation. This observation lends support to the idea that global patterns in gene content emerge from regularity in the interactions between genotypes due to conserved genotype-phenotype relationships.

Discussion

Surveys of microbiomes in contexts from hosts to oceans have elucidated environmentally-mediated patterns in the composition of these complex communities. Understanding these patterns, and ultimately their implications for the function and persistence of microbial consortia is a long-standing problem in ecology. We have presented a data-driven and empirically-validated approach to understanding how changing environmental conditions drive changing community composition at the level of gene content. Our results reveal that environmentally-mediated patterns on a global scale can emerge from interactions between members of a community. We arrived at this conclusion by first dissecting which environmental factors correlate with variation in gene abundances in the global topsoil microbiome, finding pH to be strongly correlated with the composition of the denitrification pathway. Our statistical approach motivated an experiment varying pH and assaying the structure of resulting denitrifying communities in the laboratory. Our enrichments recapitulated the observed pattern, allowing us to dissect interactions quantitatively. We showed that the dominance of the Nar^+ genotype in acidic conditions relies on the presence of a Nap$+$ genotype to reduce pH-dependent nitrite toxicity.

The central finding of our study is that genomic patterns on a global scale can emerge from ecological interactions between members of the community. The finding points to the key idea that patterns can emerge from interactions in communities as a whole, and not necessarily from specific traits conferring an advantage to individual taxa in certain environmental conditions (*45*). While filtering clearly plays a role in defining these patterns, for example at neutral pH in our study, interactions cannot be ignored.

Limitations and extensions

Our analysis of gene content in the global topsoil microbiome revealed a pattern of gene abundances across a gradient of pH from very acidic soils (pH ~ 4) to alkaline (pH ~ 8). Over this range, we observed continuous variation in the abundances of the narG and napA genes. However, our

enrichment experiments were restricted to pH 6.0 and 7.3. As a result, it remains unclear whether or not other interactions, or even filtering, might play a role in defining the full extent of the global pattern we observe. For example, it is common to find obligate nitrate reducers possessing *narG* that do not reduce nitrite (*20*). These genotypes, and any interactions that they participate in, could be essential for fully understanding the global pattern we observed. Interrogating interactions in more acidic conditions could be useful in uncovering additional processes.

Second, our statistical analysis of variation in pathway magnitude and composition focused on variations in the composition of the denitrification pathway across the globe. As a result, our study shows that interactions are important for determining which genotypes participate in the process of denitrification. However, the magnitude of the pathway, i.e., approximately the fraction of reads in the metagenome mapping to denitrification enzymes, decreased with increasing C/N ratio in soils ($\rho < 0$, Fig. S1). Denitrification is known to compete with dissimilatory nitrate reduction to ammonia (DNRA) in anaerobic environments (*28*), with low C/N favoring denitrification over DNRA for reasons of redox balance (*46*). Our findings are consistent since rising C/N inhibits denitrification globally. In this sense, it is likely that an environmental filtering framework does indeed explain which process, denitrification versus DNRA, dominates in a given environmental context. It will be an important avenue for future work to use the methods developed here to generate concrete hypotheses for the role of filtering and interactions in pattern formation. Testing these will be important for understanding the fate of nutrients in the environment, as DNRA results in the production of bioavailable ammonia, and denitrification drives nitrogen loss to N_2.

Trait conservation and trade-offs

A key finding of our study is the idea that consistent patterns in gene content on a global scale can reflect conserved genotype-phenotype relationships across diverse bacteria. This result is in agreement with our previous work showing that denitrification genotypes confer similar phenotypes across different taxa (*20*). The present study suggests that this conservation of traits can lead to

the conservation of interactions between genotypes and thus patterns observable via metagenomic sequencing over many environments.

In our system, the interaction between Nar^+ and Nap^+ genotypes arises due to specialization by these two genotypes into the first and second steps of the denitrification cascade. This specialization suggests the possibility that there is a physiological trade-off between nitrate and nitrite reduction rates. Previous studies have implicated such a trade-off and suggested that it might emerge from competition for intracellular resources (*30, 47*). We investigated this further using simulations that account for the trade-off between nitrate and nitrite reduction rates and the toxicity of nitrite, which preferentially inhibits nitrate specialists (Supplemental Information). These simulations identified two phases of community assembly: (I) with high toxicity (low pH) and concave trade-off, we observed a regime where two generalist strains that perform both nitrate and nitrite reduction coexist (Fig. S9B), and (II) when toxicity is low (neutral pH) and the front is concave, a single generalist (reduces both nitrate and nitrite) dominates (Fig. S9B). Thus, we found that the observed experimental phenomena at both pH 6.0 (regime I) and pH 7.3 (regime II) emerge generically from the imposition of empirically observed physiological constraints. Together, these findings suggest the idea that learning conserved physiological constraints on traits will be an important step in understanding ecological processes and patterns on a global scale.

Learning from environmental variation in the wild

Our results demonstrate the power of statistical analysis to motivate *in vitro* experiments that connect strain-level phenomena to macroecological patterns. Our approach used a statistical decomposition of variation in natural communities to reveal patterns in gene abundances. This approach is in contrast to a "mechanism-first" approach recently undertaken by Abreu *et al.* (*48*) that identified the phenomenon of slow growth being favored at warmer temperatures in a simplified laboratory community, before searching for this pattern in the wild. While both approaches are powerful for understanding patterns in nature, a statistical approach allows us to discover patterns and processes

without first postulating the environmental variables responsible or the details of the interaction in the laboratory. In this way, our approach allows us to uncover the key environmental drivers of variation in the wild before dissecting the role of interactions in the community.

The stress gradient hypothesis

The stress gradient hypothesis (SGH) provides a framework for understanding how the sign of ecological interactions depends on environmental conditions (*49*). The SGH suggests that interactions should skew positive (facilitation) in the presence of an environmental stressor, and negative (competitive) in the absence of stress. Previous attempts to demonstrate the SGH in the plant and microbial world highlight challenges associated with the ambiguity of the hypothesis (*50*). Namely, not all stressors appear to select for facilitative interactions (*51*), and in situations where the SGH has been observed (*52*), it is unclear whether any conserved underlying mechanisms are responsible. These challenges mean that we do not yet know whether a systematic description of the impact of stress on community interactions is plausible.

Our results represent a clear demonstration of the SGH, with nitrite toxicity at low pH serving as the stressor and the alleviation of toxicity by Nap^+ genotypes serving as the facilitative interaction. More importantly, our approach demonstrates a systematic methodology for elucidating the SGH in communities. Specifically, we have shown that conserved genotype-phenotype relationships can drive patterns of facilitative interactions, leaving a signature of the SGH in covariation between environmental variables and metagenomic data. Further analyses of these datasets could yield insights into the importance of the SGH for other microbial processes. Such analyses could enable a more systematic interrogation of the relevance of the SGH in structuring natural communities.

The environmental filtering hypothesis

The concept of environmental filtering has a long history, perhaps beginning with the foundational work of von Humbolt and Bonpland (*53*), and the belief that traits alone can determine community composition has remained persistent when interpreting data in the modern era. Surveys in marine

microbiomes show that functional traits correlate with environmental variables (*54*). Similarly, studies of the gut microbiome have interpreted reproducible changes in the abundances of specific taxa in response to diet in terms of environmental filtering (*55*). In these very complex ecological contexts, it is challenging to definitively determine when interactions matter and when they do not. Experimental and analytical methods for dissecting interactions in these contexts will be important going forward.

More broadly, we hope this work helps establish a link between two conceptual frameworks in community ecology. First, decades of work have focused on the structure of interactions between taxa and the resulting abundance dynamics (*56*). In this picture, traits are often assigned randomly, and community function or ecosystem services play a secondary role. Second, quantitative physiology and an appreciation of the importance of metabolic traits in community assembly (*57*) have revealed that traits are not random, but rather that strong conservation yields regular structure in the physiology of microbes, yielding robust quantitative principles (e.g., Ref. 58). Our study proposes a path forward that bridges these two conceptual frameworks, revealing how conserved traits can give rise to recurrent patterns of interactions in wild communities with implications for ecosystem functions. We hope that understanding the origins of these patterns will drive deeper insights into the assembly and function of complex natural microbiomes.

Acknowledgments

We acknowledge Luis Miguel de Jesús Astacio and Kiseok Lee for collecting soils, Kiseok Lee for help with DNA library preparation, Mohammad Bahram for previously published global topsoil microbiome data, and Kabir Husain for useful discussions.

Funding: This work was supported by the National Science Foundation Division of Emerging Frontiers EF 2025293 (S.K.) and EF 2025521 (M.M.), as well as, the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1746045 (M.S.C-W). S.K. acknowl-

edges the Center for the Physics of Evolving Systems at the University of Chicago. K.G. acknowledges a James S. McDonnell Foundation Postdoctoral Fellowship Award 220020499. M.S.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., and M.C-W. performed the experiments. K.G. performed statistical analysis of the topsoil microbiome data and performed and analyzed soil enrichment experiments. M.C-W. performed mortality rate experiments, advised by K.C. and S.K.. Z.L. performed co-evolution analysis, advised by K.G. and S.K.. K.C. performed all other experiments and analyses, advised by K.G. and S.K.. K.C. performed CRM fits, predictions, and simulations, adapting code written by K.G.. K.C., K.G., Z.L., and S.K. wrote the manuscript with input from all authors.

Competing interests: The authors declare no competing interests.

Data and materials availability: Code and data associated with this manuscript are publicly available at `https://github.com/kyle-crocker/publicationCodeData/tree/main/interactionPatterns` and by request. Raw sequence reads for enrichment experiments are deposited under NCBI BioProject ID PRJNA976277. Bacterial isolates are available by request.

References and Notes

1. P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034 (2008).

2. W. Song, *et al.*, *mBio* **13**, e03832 (2022).

3. C.-y. Wang, *et al.*, *Annals of Microbiology* **69**, 1461 (2019).

4. M. Bahram, *et al.*, *Nature* **560**, 233 (2018).

5. A. A. Malik, B. C. Thomson, A. S. Whiteley, M. Bailey, R. I. Griffiths, *mBio* **8**, e00799 (2017).

6. M. H. Sofi, *et al.*, *Diabetes* **63**, 632 (2014).

7. M. K. Zinöcker, I. A. Lindseth, *Nutrients* **10**, 365 (2018).

8. R. K. Singh, *et al.*, *Journal of Translational Medicine* **15**, 1 (2017).

9. C. L. Lauber, M. Hamady, R. Knight, N. Fierer, *Applied and Environmental Microbiology* **75**, 5111 (2009).

10. J. Rousk, *et al.*, *The ISME Journal* **4**, 1340 (2010).

11. P. A. Dimitriu, S. J. Grayston, *Microbial Ecology* **59**, 563 (2010).

12. R. T. Jones, *et al.*, *The ISME Journal* **3**, 442 (2009).

13. K. Faust, J. Raes, *Nature Reviews Microbiology* **10**, 538 (2012).

14. A. R. Pacheco, D. Segrè, *FEMS Microbiology Letters* **366**, fnz125 (2019).

15. R. H. Hsu, *et al.*, *Cell Systems* **9**, 229 (2019).

16. E. B. Sherr, B. F. Sherr, *Antonie van Leeuwenhoek* **81**, 293 (2002).

17. M. de la Cruz Barron, *et al.*, *The ISME Journal* pp. 1–11 (2023).

18. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, *Current Biology* **29**, R521 (2019).

19. O. S. Venturelli, *et al.*, *Molecular Systems Biology* **14**, e8157 (2018).

20. K. Gowda, D. Ping, M. Mani, S. Kuehn, *Cell* **185**, 530 (2022).

21. W. G. Zumft, *Microbiology and Molecular Biology Reviews* **61**, 533 (1997).

22. D. A. Rodionov, I. L. Dubchak, A. P. Arkin, E. J. Alm, M. S. Gelfand, *PLoS computational biology* **1** (2005).

23. K. Heylen, *et al.*, *Applied and Environmental Microbiology* **72**, 2637 (2006).

24. P. Lycus, *et al.*, *ISME Journal* **11**, 2219 (2017).

25. T. N. Gamble, M. R. Betlach, J. M. Tiedje, *Applied and Environmental Microbiology* **33**, 926 (1977).

26. J. P. Shapleigh, *The Prokaryotes* (2006), pp. 769–792.

27. J. Uhlmann (2016).

28. J. M. Tiedje, A. J. Sexstone, D. D. Myrold, J. A. Robinson, *Antonie van Leeuwenhoek* **48**, 569 (1983).

29. M. Šimek, J. E. Cooper, *European Journal of Soil Science* **53**, 345 (2002).

30. E. E. Lilja, D. R. Johnson, *ISME Journal* **10**, 1568 (2016).

31. F. Magdoff, R. Bartlett, *Soil Science Society of America Journal* **49**, 145 (1985).

32. P. N. Nelson, N. Su, *Soil Research* **48**, 201 (2010).

33. B. Bache, *Plant, Cell & Environment* **7**, 391 (1984).

34. C. W. Kreitler, D. C. Jones, *Groundwater* **13**, 53 (1975).

35. R. J. Dowdell, R. Q. Cannell, *Journal of Soil Science* **26** (1975).

36. R. Logares, *et al.*, *Environmental Microbiology* **16**, 2659 (2014).

37. J. R. Cole, *et al.*, *Nucleic Acids Research* **33**, D294 (2005).

38. F. Meyer, *et al.*, *BMC Bioinformatics* **9**, 1 (2008).

39. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, *Applied and Environmental Microbiology* **73**, 5261 (2007).

40. G. V. Uritskiy, J. Diruggiero, J. Taylor, *Microbiome* **6**, 1 (2018).

41. J.-M. Bollag, N. M. Henninger, *Soil Biology and Biochemistry* **10**, 377 (1978).

42. H. Tarr, *Journal of the Fisheries Board of Canada* **5**, 265 (1941).

43. K. Bancroft, I. F. Grant, M. Alexander, *Applied and Environmental Microbiology* **38**, 940 (1979).

44. K. Durvasula, *et al.*, *Biotechnology progress* **25**, 973 (2009).

45. N. J. Kraft, *et al.*, *Functional Ecology* **29**, 592 (2015).

46. E. M. van den Berg, U. van Dongen, B. Abbas, M. C. van Loosdrecht, *The ISME Journal* **9**, 2153 (2015). Number: 10 Publisher: Nature Publishing Group.

47. M. Rigoulet, A. Mourier, A. Galinier, L. Casteilla, A. Devin, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1797**, 671 (2010).

48. C. I. Abreu, M. Dal Bello, C. Bunse, J. Pinhassi, J. Gore, *Science Advances* **9**, eade8352 (2023). Publisher: American Association for the Advancement of Science.

49. M. D. Bertness, R. Callaway, *Trends in Ecology and Evolution* **9**, 187 (1994).

50. S. P. Hammarlund, W. R. Harcombe, *Proceedings of the National Academy of Sciences of the United States of America* **116**, 15760 (2019).

51. F. T. Maestre, F. Valladares, J. F. Reynolds, *Journal of Ecology* **93**, 748 (2005).

52. P. Piccardi, B. Vessman, S. Mitri, *Proceedings of the National Academy of Sciences of the United States of America* **116**, 15979 (2019).

53. A. v. Humboldt, A. Bonpland, *Essay on the Geography of Plants* (University of Chicago Press, Chicago, 1838).

54. S. Louca, L. W. Parfrey, M. Doebeli, *Science* **353**, 1272 (2016).

55. L. A. David, *et al.*, *Nature* **505** (2014).

56. R. M. May, *Stability and Complexity in Model Ecosystems*, vol. 1 (Princeton University Press, 1974).

57. J. B. H. Martiny, S. E. Jones, J. T. Lennon, A. C. Martiny, *Science* **350** (2015).

58. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, *Science* **330**, 1099 (2010).

59. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

60. B. Bushnell, J. Rood, E. Singer, *PloS One* **12** (2017).

61. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, *Genome Research* **27**, 824 (2017).

62. J. Alneberg, *et al.*, *Nature Methods* **11**, 1144 (2014).

63. Y.-W. Wu, B. A. Simmons, S. W. Singer, *Bioinformatics* **32**, 605 (2016).

64. D. D. Kang, J. Froula, R. Egan, Z. Wang, *PeerJ* **2015**, 1 (2015).

65. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Research* **25**, 1043 (2015).

66. C. Quast, *et al.*, *Nucleic Acids Research* **41**, D590 (2012).

67. F. O. Glöckner, *et al.*, *Journal of Biotechnology* **261**, 169 (2017).

68. P. Yilmaz, *et al.*, *Nucleic Acids Research* **42**, D643 (2014).

69. T. Waki, K. Murayama, Y. Kawato, K. Ichikawa, *Journal of Fermentation Technology* **58** (1980).

70. P.-H. Liu, G. Zhan, S. A. Svoronos, B. Koopman, *Water Research* **32**, 3452 (1998).

71. P.-H. Liu, S. A. Svoronos, B. Koopman, *Biotechnology and Bioengineering* **60**, 649 (1998).

72. M. Gouw, R. Bozic, B. Koopman, S. A. Svoronos, *Water Research* **35**, 2794 (2001).

73. K. Lisbon, M. McKean, S. Shekar, S. A. Svoronos, B. Koopman, *Journal of Environmental Engineering* **128**, 391 (2002).

74. A. I. Casasús, R. K. Hamilton, S. A. Svoronos, B. Koopman, *Water Research* **39**, 1914 (2005).

75. 16s metagenomic sequencing library preparation, *Tech. rep.*, Illumina, Inc., San Diego, CA, USA (2013).

76. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, *Bioinformatics* **30**, 614 (2014).

77. B. J. Callahan, *et al.*, *Nature Methods* **13**, 581 (2016).

78. N. A. Bokulich, *et al.*, *Microbiome* **6**, 1 (2018).

79. M. R. McLaren, A. D. Willis, B. J. Callahan, *Elife* **8**, e46923 (2019).

80. H. H. Ku, *et al.*, *Journal of Research of the National Bureau of Standards* **70** (1966).

81. J. R. Taylor, *An Introduction to Error Analysis (Second Edition)* (University Science Books, 1997).

82. M. Kanehisa, S. Goto, *Nucleic Acids Research* **28**, 27 (2000).

83. M. Kanehisa, *Protein Science* **28**, 1947 (2019).

84. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, *Nucleic Acids Research* **49**, gkaa970 (2020).

85. Y. Matsumoto, *et al.*, *Nature Structural and Molecular Biology* **19**, 238 (2012).

86. A. Strehl, J. Ghosh, *Journal of Machine Learning Research* **3**, 583 (2002).

87. A. F. Schober, *et al.*, *Cell Reports* **27**, 3359 (2019).

88. N. I. van den Berg, *et al.*, *Nature Ecology & Evolution* **6**, 855 (2022).

89. K. M. Miranda, M. G. Espey, D. A. Wink, *Nitric Oxide - Biology and Chemistry* **5**, 62 (2001).