

RESEARCH ARTICLE

Distinct temporal diversity profiles for nitrogen cycling genes in a hyporheic microbiome

William C. Nelson^{1*}, Emily B. Graham¹, Alex R. Crump², Sarah J. Fansler¹, Evan V. Arntzen¹, David W. Kennedy¹, James C. Stegen¹

1 Pacific Northwest National Laboratory, Richland, Washington, United States of America, **2** Department of Soil and Water Systems, University of Idaho, Moscow, Idaho, United States of America

* william.nelson@pnnl.gov



OPEN ACCESS

Citation: Nelson WC, Graham EB, Crump AR, Fansler SJ, Arntzen EV, Kennedy DW, et al. (2020) Distinct temporal diversity profiles for nitrogen cycling genes in a hyporheic microbiome. PLoS ONE 15(1): e0228165. <https://doi.org/10.1371/journal.pone.0228165>

Editor: Wenzhi Liu, Chinese Academy of Sciences, CHINA

Received: October 21, 2019

Accepted: January 8, 2020

Published: January 27, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0228165>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Data sets are available through the JGI Genome Portal (<http://genome.jgi.doe.gov>). Project identifiers are listed in [S1 Table](#) in the Supporting Information file.

Abstract

Biodiversity is thought to prevent decline in community function in response to changing environmental conditions through replacement of organisms with similar functional capacity but different optimal growth characteristics. We examined how this concept translates to the within-gene level by exploring seasonal dynamics of within-gene diversity for genes involved in nitrogen cycling in hyporheic zone communities. Nitrification genes displayed low richness—defined as the number of unique within-gene phylotypes—across seasons. Conversely, denitrification genes varied in both richness and the degree to which phylotypes were recruited or lost. These results demonstrate that there is not a universal mechanism for maintaining community functional potential for nitrogen cycling activities, even across seasonal environmental shifts to which communities would be expected to be well adapted. As such, extreme environmental changes could have very different effects on the stability of the different nitrogen cycle activities. These outcomes suggest a need to modify existing conceptual models that link biodiversity to microbiome function to incorporate within-gene diversity. Specifically, we suggest an expanded conceptualization that 1) recognizes component steps (genes) with low diversity as potential bottlenecks influencing pathway-level function, and 2) includes variation in both the number of entities (e.g. species, phylotypes) that can contribute to a given process and the turnover of those entities in response to shifting conditions. Building these concepts into process-based ecosystem models represents an exciting opportunity to connect within-gene-scale ecological dynamics to ecosystem-scale services.

Introduction

High microbial diversity has been observed in almost all environments that have been examined [1]. It is widely believed that this diversity provides functional stability to ecosystems experiencing fluctuations in environmental conditions by the presence of organisms having overlapping functional capabilities but different conditions under which they optimally

Funding: This research and all authors were supported by the US Department of Energy, Office of Biological and environmental Research (<https://www.energy.gov/science/ber/biological-and-environmental-research>), as part of Subsurface Biogeochemical Research Program's (<https://www.doesbr.org>) Scientific Focus Area (SFA) at the Pacific Northwest National Laboratory (PNNL). PNNL is operated for DOE by Battelle under contract DE-AC06-76RLO 1830. Sequencing was done at the DOE Joint Genome Institute under Community Science Project 1781 awarded to J.C. S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

function [2–8]. In a fluctuating environment, conditions that impair the growth of some populations will stimulate the growth of others, and overall community function is maintained. Maintenance of higher diversity therefore allows a community to respond more rapidly to a disturbance or environmental shift and reduces its dependence on (or susceptibility to) recruitment of new organisms to fill vacant niches. The dynamics of diversity at the functional gene level, however, have not been well explored.

Cooperative metabolism in natural microbial communities has long been suspected, but only recently have metagenomic studies revealed its extent. The component steps (*i.e.*, individual enzyme-catalyzed reactions) of complex metabolic pathways, such as denitrification, sulfur oxidation, and organic carbon degradation, have been observed to be distributed across multiple organisms more frequently than they are co-resident in a single organism [9, 10]. Distributed metabolism likely reflects efficiency gains from specialization and division of labor [11]. This partitioning, however, puts component steps of critical ecosystem processes under different selective pressures, according to which organism encodes them. Temporal dynamics of diversity and abundance may, therefore, vary significantly across component steps.

Nitrogen cycling is an excellent and ubiquitous example of a complex, distributed process. A generalized model of the dominant N processes in the HZ (Fig 1) includes conversion of NH_4^+ to NO_3^- (*i.e.*, *nitrification*) in oxic regions, which is coupled to carbon (C) fixation, and reduction of NO_3^- coupled to the oxidation of organic carbon (OC) in anoxic regions. This latter process of *denitrification* yields N_2 gas, removing N from the system. Nitrite (NO_2^-) is an intermediate common to both processes. While complete denitrifier organisms, such as *Pseudomonas aeruginosa* and *Parcoccus denitrificans*, have been isolated and described, it has long been suspected that many organisms encode partial pathways and can act in concert to cycle nitrogen between its reduced and oxidized forms [12]. More recently, genome sequence data from both isolates and environmental samples has shown that many organisms encode various subsets of denitrification activities [9, 13]. Several previous studies have investigated the abundance and distribution of nitrogen cycling activities in environmental microbiomes [14–19], none yet have specifically tracked the diversity of individual gene families that comprise nitrogen transformation pathways across fluctuating environmental conditions.

Here we take advantage of seasonal shifts in hydrology and aqueous geochemistry within a hyporheic zone system that have been shown to alter microbial community structure [20, 21], and examine the temporal dynamics of diversity within major N-cycling genes encoding steps in nitrification and denitrification. Some component steps consistently showed very low diversity, while others displayed significant temporal variation in the level of diversity and turnover in the contributing phylotypes across divergent environmental conditions. The observed heterogeneity through time and across component steps indicates that predictive ecosystem models that explicitly represent microbial communities should account for variation in and dynamics of within-gene diversity of component steps of key processes.

Results

Seasonal environmental changes

Sediment communities from the hyporheic zone of the Columbia River along the Hanford Reach were sampled from April 30, 2014 to November 25, 2014, using sand packs deployed at three equivalent hyporheic zone locations approximately 100m apart along the river (T2, T3, and T4) for six weeks at a time [22]. Water chemistry data taken in parallel at the three sites showed similar, yet not identical temporal patterns. A mid-year shift in hydraulic regime was observed, with higher influx of surface water in the spring resulting in higher levels of dissolved organic carbon (measured as non-purgeable organic carbon) (NPOC) (0.8–1.0 mg/L)

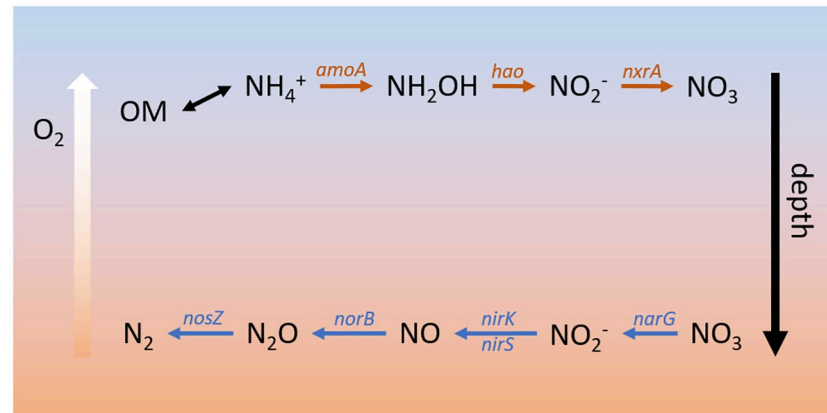


Fig 1. Major nitrogen transformations in the hyporheic zone (HZ). Upper layers (closer to the surface channel) of the hyporheic zone contain more oxygen (O_2) and organic matter (OM). Under these conditions, *nitrification* (orange arrows) occurs. Ammonium (NH_4^+) is released from by OM breakdown and converted to nitrate (NO_3^-) through hydroxylamine and nitrite (NO_2^-) intermediates. This process has been linked with carbon fixation, increasing organic carbon (OC). Aerobic respiration depletes O_2 , causing deeper regions of the HZ to become hypoxic or anaerobic. Under these conditions, *denitrification* (blue arrows) converts nitrate to nitrogen gas (N_2) through nitrite, nitric oxide (NO) and nitrous oxide intermediates, and provides an electron acceptor for catabolism of OM.

<https://doi.org/10.1371/journal.pone.0228165.g001>

(Fig 2A) and low levels of nitrate (10–15 μM) (Fig 2B), transitioning to a more groundwater-influenced condition in the fall, increasing the nitrate concentrations (up to 300 μM) and decreasing NPOC concentration (down to <0.4 mg/L). Because the groundwater in this system is oxic, the DO concentration was fairly constant for the duration of sampling, ranging from ~60–100% saturation (Fig 2C). The water temperature followed expected seasonal trends, warming in the summer and cooling in the fall (Fig 2D). Sampling times were categorized as early (Apr 30 through Jul 22) or late (Sep 2 through Nov 30), based on these observations.

Organism-level diversity

Organismal diversity was measured by 16S rRNA V4 amplicon sequence analysis and extraction and assembly of *rplB* gene sequences from the metagenomic data sets (Fig 3). As reported previously [20], species richness correlated best with water temperature. Diversity, as measured by the inverse Simpson statistic, was high and mirrored species richness, suggesting high evenness. Two late samples, October 14 and November 25, showed high richness but low diversity, driven by a bloom of Bacteroidetes species.

Diversity of N-cycling genes

The temporal phylogenetic profile of each gene of interest was examined to elucidate the richness and diversity of genes comprising the nitrification and denitrification processes. Metagenomic reads containing sequence from the genes of interest were extracted from the total data set and assembled to yield partial and full-length gene sequences (Supplementary Data 1). Phylogeny was determined for each assembled sequence, and phylotypes were defined at 90% amino acid sequence identity, since that level of similarity is typical between organisms of the same genus [23]. Richness was quantified for each gene as the number of distinct phylotypes identified. It was expected that detectable gene diversity would be considerably lower than organismal diversity, since 1) these activities are encoded by a subset of organisms, and 2) the assembly protocol is less sensitive than amplicon analysis, and thus only genes from abundant organisms are likely to be detected. The relative abundance of each phylotype was estimated

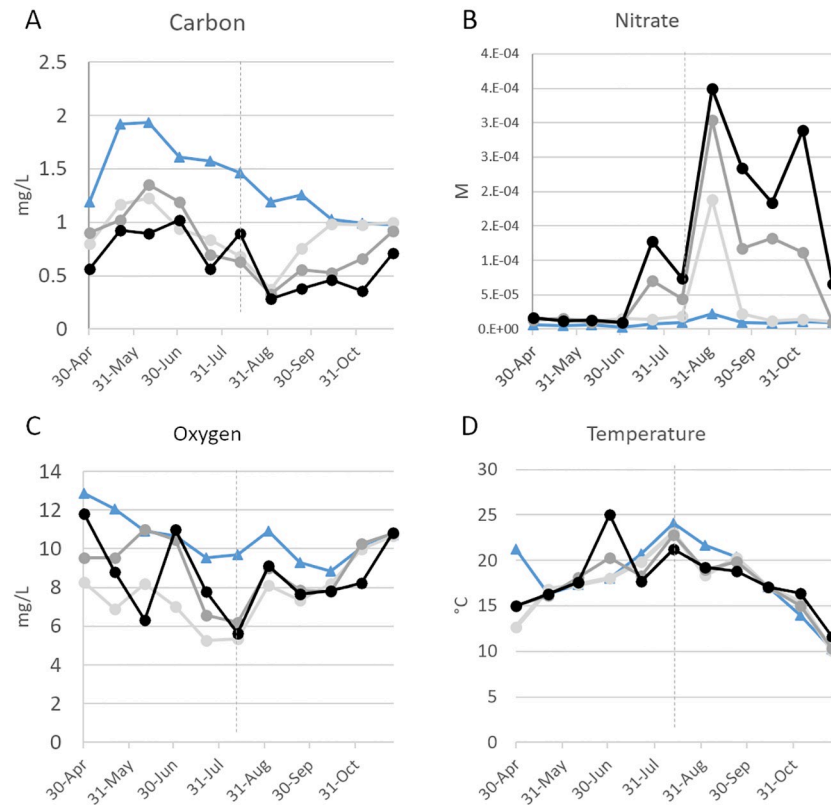


Fig 2. Water chemistry and temperature of sampled sites. Piezometer T2, light gray; piezometer T3, dark gray; piezometer T4, black. For comparison, data for adjacent river water is presented (blue). The vertical dotted line indicates the date at which the hyporeheic zone hydraulic regime changes from surface water intrusion to groundwater discharge.

<https://doi.org/10.1371/journal.pone.0228165.g002>

from the summed assembly coverage of the member genes. Temporal diversity dynamics (turnover) were assessed by calculating the mean variance of relative abundance for phylotypes across time.

Distinct diversity and turnover patterns were observed for each gene. The *narG* and *nosZ* genes (Figs 4 and 5, summarized in Table 1), encoding the first and last steps of the denitrification process, respectively, had higher phylotype richness than the other nitrogen cycle genes examined (for *nosZ* vs *norB*, Welch’s t-test p-value = 0.0014, df = 13.587), and their phylotype profiles had equivalent stability (Levene test p-value = 0.1277). While the *nirK/nirS* (distinct types of nitrite reductase) (Fig 6) and *norB* family (nitric oxide reductase) (Fig 7) had lower richness, their phylotype profile variability was significantly higher than for *narG* (Levene test p-values = 0.00003, 0.0001, respectively), and were near significance for *nosZ* (Levene test p-values = 0.0113, 0.0609 for *nosZI* and *nosZII*). Both genes encoding activities involved in nitrification had extremely low phylotype diversity, *amoA* (ammonia oxidase) with 2 phylotypes, one bacterial and one archaeal, and *nxrA* (nitrite oxidase alpha subunit) having 7 observed, but one overwhelmingly dominant phylotype (Fig 8A). The low richness for *amoA* (Fig 8B) exaggerates the phylotype abundance variance values, thus we consider the low richness to be the significant aspect of the *amoA* gene.

Nitrogen gene diversity was largely dependent upon a temporally consistent pool of taxa. An examination of cumulative phylotype richness (Fig 9A) showed an increase in the number phylotypes detected for almost all the target gene families in the spring, with limited increase

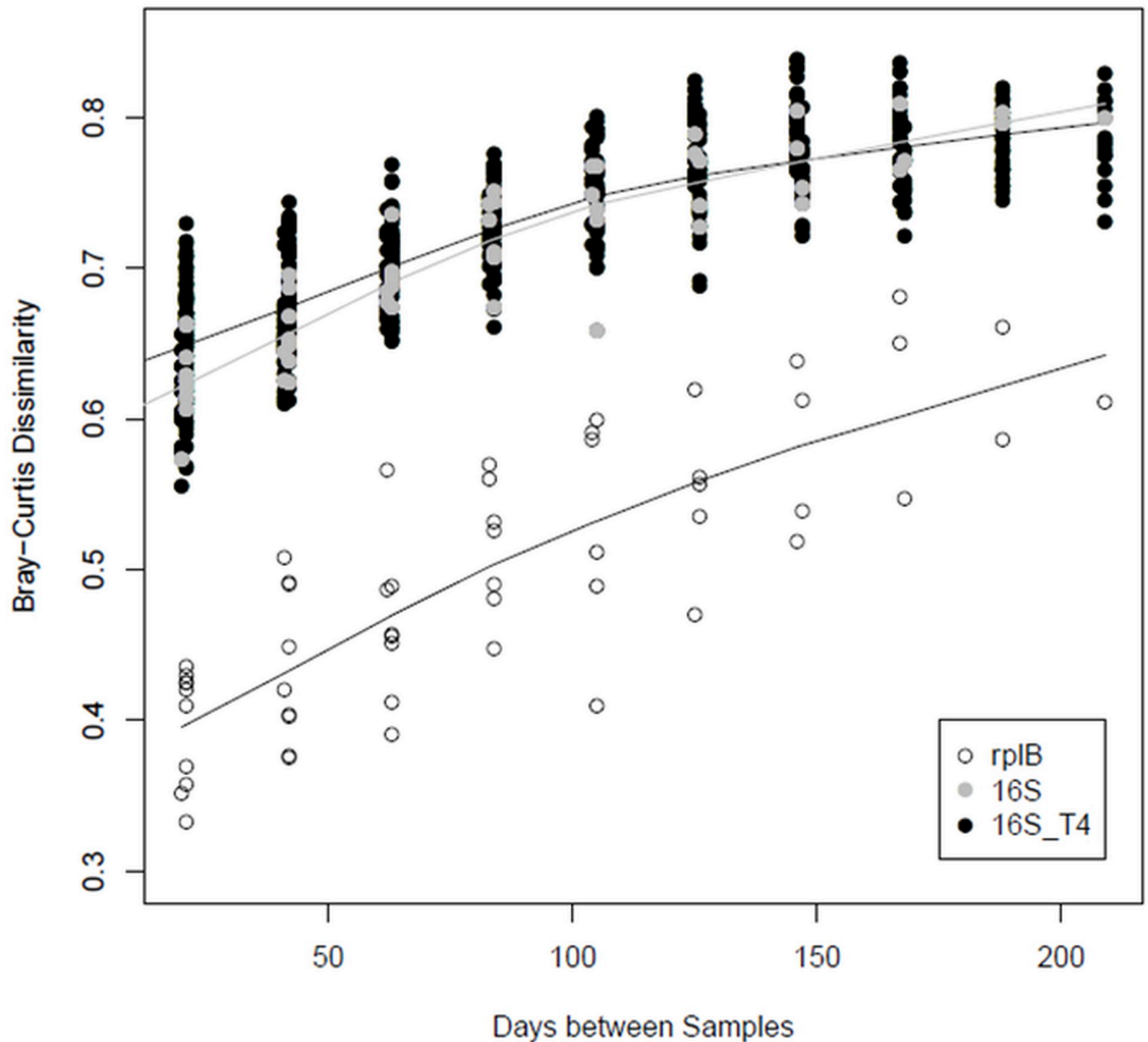


Fig 3. Sediment microbial community continuously changes across the year. Distance-decay plot of all 16S rRNA amplicon data ("16S"), amplicon data from only site T4 ("16S_T4"), and *rplB* genes ("rplB") assembled from all metagenomes.

<https://doi.org/10.1371/journal.pone.0228165.g003>

thereafter. Importantly, cumulative richness curves generated using unique sequences, rather than phylotypes, have equivalent shape (data not shown). An analysis of cumulative diversity (inverse Simpson) showed that the increase in the number of phylotypes had a proportional effect on diversity except in the cases of *narG*, where the large increase in phylotypes only translates to a modest increase in diversity, and *nxrA* where the small increase in richness had no effect on diversity (Fig 9B). This was due to the additional phylotypes having low relative abundance. The *nirKS* family showed an initial decrease in diversity despite an increase in the number of phylotypes, and a subsequent increase in diversity with no further increase in richness. This increase in diversity is due to increasing evenness amongst the various phylotypes present. The *nosZ* family was the only one to demonstrate consistent increases in diversity due to introduction of new phylotypes.

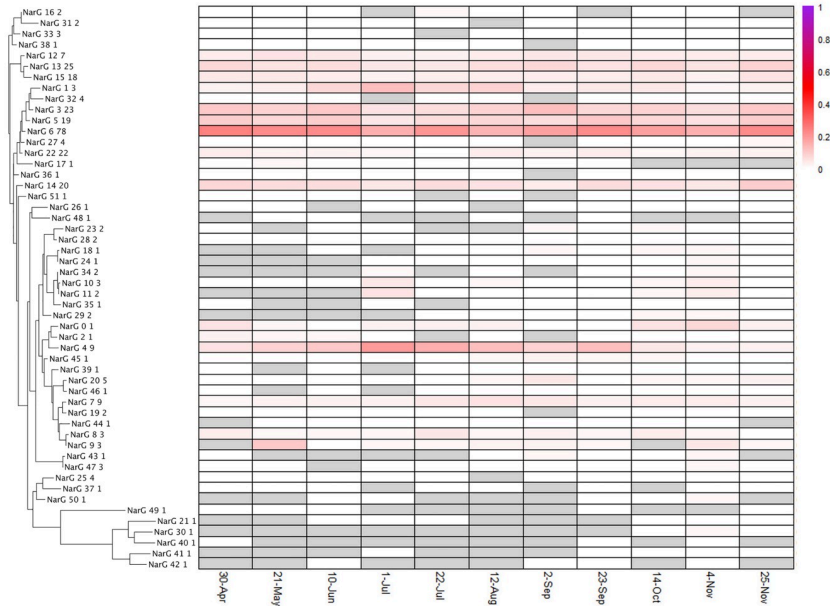


Fig 4. NarG phylotype distributions. Heatmap indicates the relative contribution of each phylotype (clustered at 90% AAID) to the total count; phylogenetic tree to the left of the heatmap demonstrates the diversity of phylotypes present. Gray shading indicates no observation of the phylotype at that timepoint.

<https://doi.org/10.1371/journal.pone.0228165.g004>

Abundance of N-cycling genes

To assess temporal changes in the overall abundances of genes involved in denitrification and nitrification, the sets of all (*i.e.*, unassembled) metagenomic reads containing sequence from the genes of interest were enumerated, and the representation of each gene within the

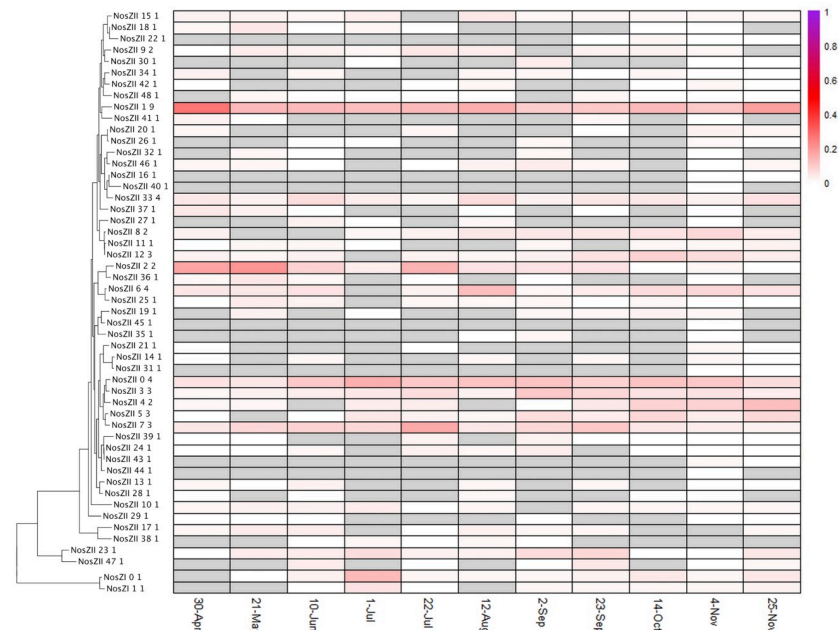


Fig 5. NosZI and NosZII phylotype distributions. See Fig 4 for description of display.

<https://doi.org/10.1371/journal.pone.0228165.g005>

Table 1. Observed richness and abundance variance.

Gene family	Phylotype richness	Mean abundance variance
<i>narG</i>	52	0.000183
<i>norB</i>	31	0.000624
<i>nirKS</i>	23	0.001095
<i>nosZ</i>	51	0.000395
<i>nxrA</i>	7	0.000047
<i>amoA</i>	2	0.089783
<i>rplB</i>	124	0.000112

<https://doi.org/10.1371/journal.pone.0228165.t001>

community was normalized across samples using counts of the conserved, single-copy *rplB* gene as a proxy for number of individuals sampled. Although gene abundances were relatively constant over time, the average abundances differed widely between genes. The *narG* gene, the first step in denitrification, was observed to be in 25–30% of the population, while the *nirK/nirS* was represented in 35–45% of the population, and *norB* in 14–18% (Fig 10). Nitrous oxide reductase genes were present in ~25% of the populations, however it is of note that the dominant form was *nosZII* (also referred to in the literature as the ‘atypical *nosZ*’), a distinct family of nitrous oxide reductases typically found in non-denitrifying organisms [13, 24, 25]. Nitrification genes showed more of a seasonal shift in abundance. The *amoA* gene, summing both

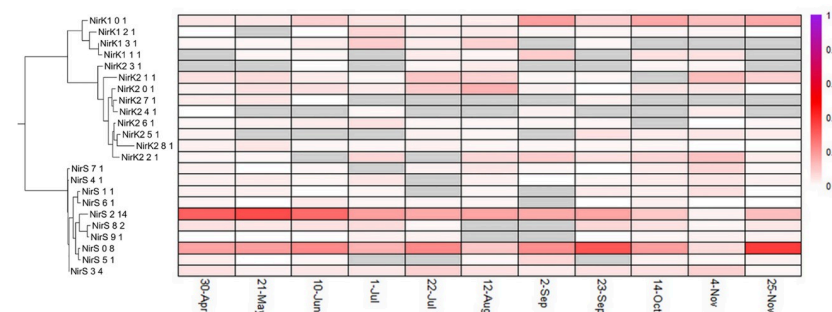


Fig 6. NirK and NirS phylotype distributions. See Fig 4 for description of display.

<https://doi.org/10.1371/journal.pone.0228165.g006>

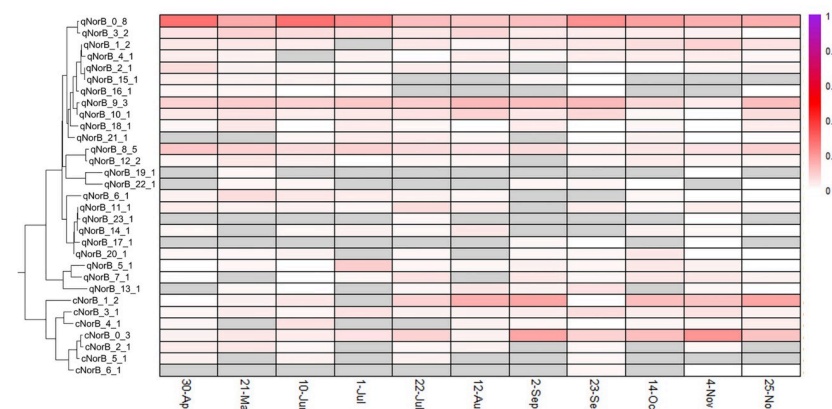


Fig 7. NorB phylotype distributions. See Fig 4 for description of display.

<https://doi.org/10.1371/journal.pone.0228165.g007>

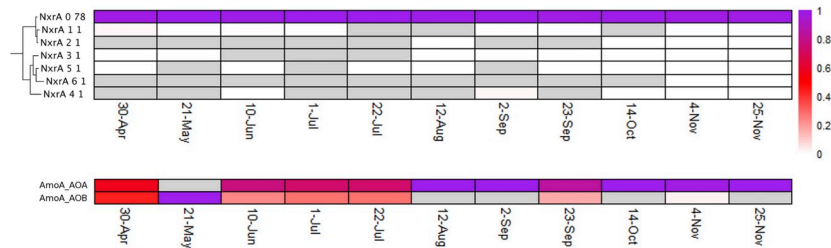


Fig 8. A) NxrA and B) AmmA phylotype distributions. See Fig 4 for description of display.

<https://doi.org/10.1371/journal.pone.0228165.g008>

the bacterial and archaeal versions, showed a low constant abundance of ~5% in early time points, and increased up near 30% late in the year. Unexpectedly, *nxrA* showed little correlation with *amoA*, displaying a trend of gradual increase, ranging from 5% to 18%, early, and constancy late.

Environmental drivers

Regression analysis was performed to determine which, if any, of the environmental parameters measured was associated with changes in diversity for the genes of interest. Water temperature, dissolved oxygen (DO), dissolved organic carbon (measured as non-purgeable organic carbon, NPOC), and chloride (Cl⁻) measurements were used. Cl⁻ is a conservative indicator of the ratio of surface- to groundwater content in the hyporheic zone of the study system [26]. Other measured constituents, NO₃⁻ and SO₄⁻ had strong positive correlations with Cl⁻ (S1 Fig). Correlations between diversity (inverse Simpson), richness, and abundance were tested

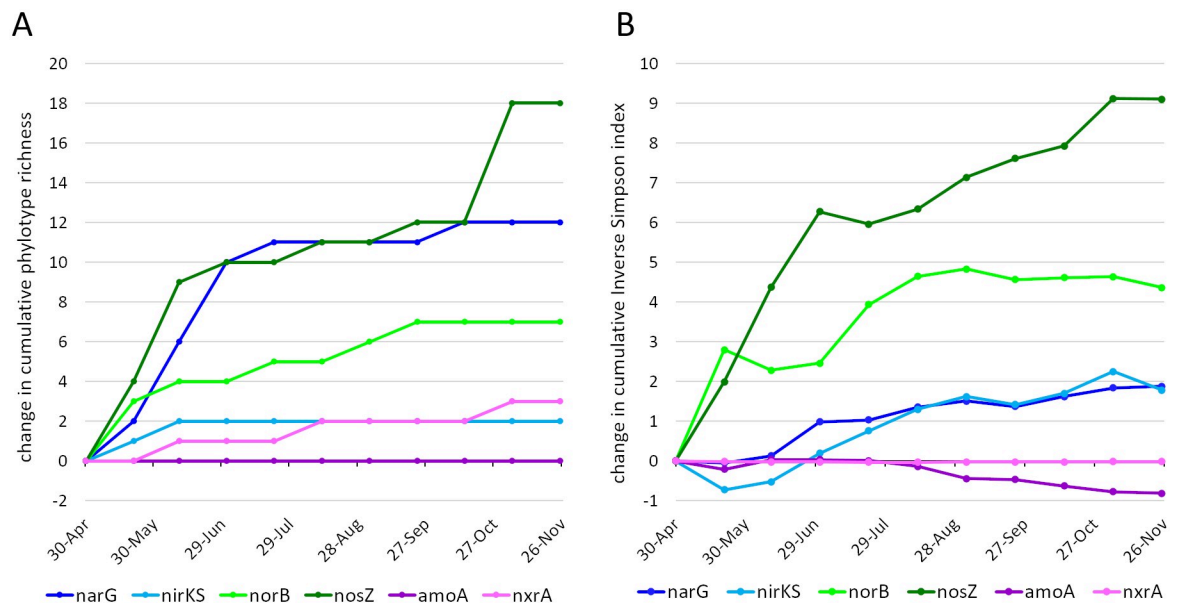


Fig 9. Cumulative diversity measures over time. Phylotype richness (A) and the inverse Simpson statistic (B) were calculated cumulatively (*i.e.*, combining data from each time point with all previous timepoints) for each gene or functional gene class (*nirK* and *nirS* counts were combined; archaeal and bacterial *amoA* types were combined). The data is presented as the difference from the initial (April 30) state. Most genes' richness values plateau, indicating sample-to-sample changes in diversity are within a finite pool of phylotypes. Diversity increase indicates either introduction of new phylotypes or increases in evenness across existing phylotypes. The decreases in diversity observed for *amoA* and *nirKS* are driven by changes in relative abundance resulting in a decrease in evenness.

<https://doi.org/10.1371/journal.pone.0228165.g009>

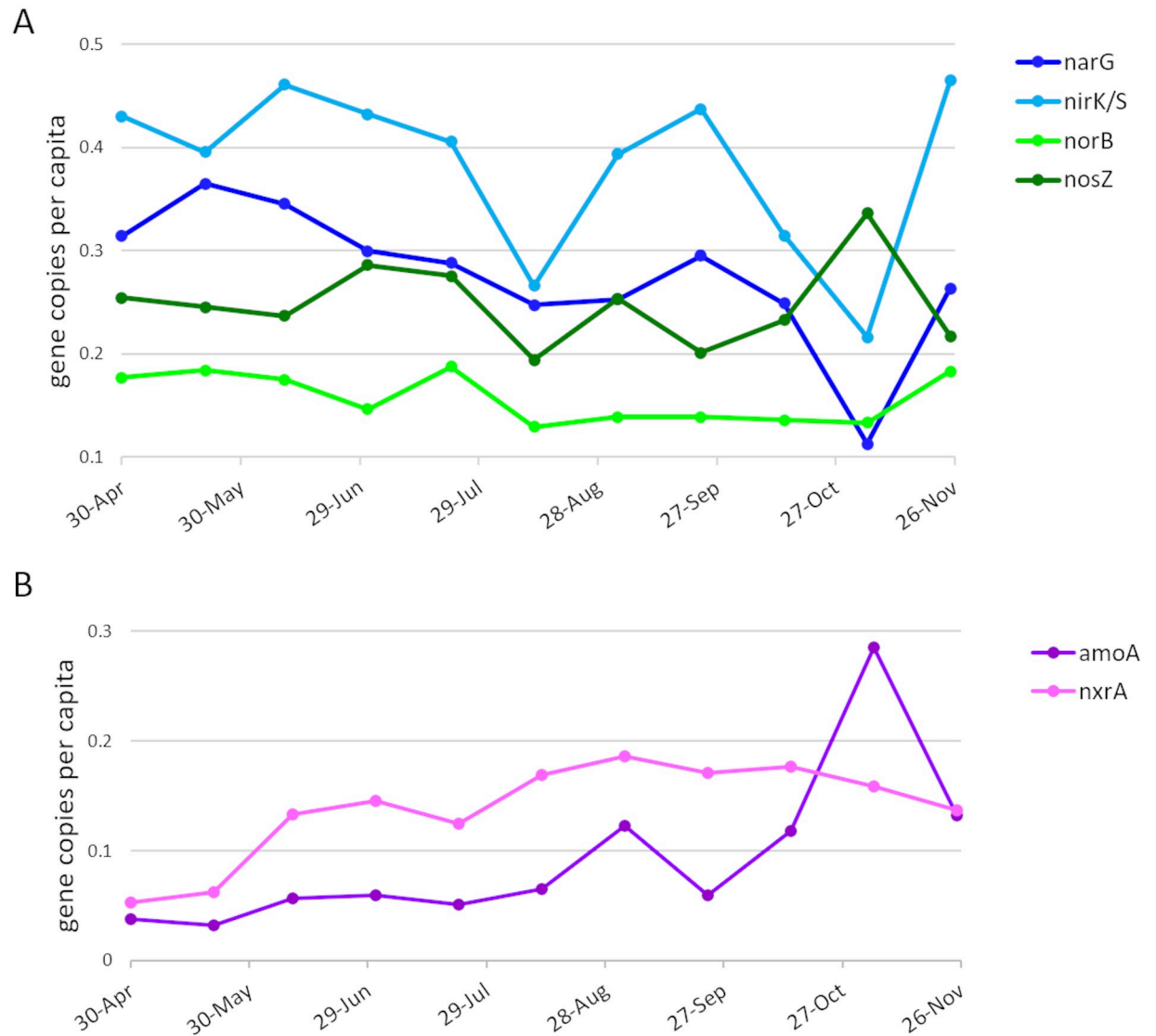


Fig 10. Per-capita abundance of denitrification and nitrification genes. Reads per kilobase of gene length per million reads (RPKM) for each gene was normalized against the RPKM for the *rplB* gene as a proxy for the number of individuals sampled. (A) Denitrification genes. (B) Nitrification genes.

<https://doi.org/10.1371/journal.pone.0228165.g010>

against the environmental parameters. The strongest relationships were with groundwater content (using Cl⁻ as a proxy), with denitrification genes *narG* ($R^2 = 0.38$; $p = 0.04$) and *nosZ* ($R^2 = 0.50$; $p = 0.02$) increasing in diversity (S2 Fig), nitrification genes *amoA* ($R^2 = 0.41$; $p = 0.03$) and *nxrA* ($R^2 = 0.44$; $p = 0.03$) increasing in abundance (S3 Fig), and *narG* ($R^2 = 0.47$; $p = 0.02$) decreasing in abundance. Groundwater showed weaker correspondence with increasing richness of *nxrA* ($R^2 = 0.29$; $p = 0.09$), decreasing richness of *nirKS* ($R^2 = 0.27$; $p = 0.10$) (S4 Fig), and decreasing abundance of *norB* ($R^2 = 0.35$; $p = 0.06$). NPOC had strongest correlations with the nitrification genes, showing a negative relationship with *nxrA* diversity ($R^2 = 0.31$; $p = 0.08$), and a positive relationship with *nirKS* richness ($R^2 = 0.39$; $p = 0.04$) and *narG* abundance ($R^2 = 0.30$; $p = 0.08$). Temperature had a significant negative relationship with *nxrA* diversity ($R^2 = 0.33$; $p = 0.08$) and richness ($R^2 = 0.45$; $p = 0.03$).

Discussion

Shade *et al.*, in their review of microbial resistance and resilience, suggest that there is “no ‘one-size fits all’ response of microbial diversity and function to disturbance.” [8]. While this perspective is undoubtedly true, it leaves open the possibility that there are general patterns or rules that govern particular subsets or components of microbial communities. Here we begin to look for such patterns at a deeper level than previously examined by exploring dynamics in gene abundance and diversity within important biogeochemical processes in response to seasonal environmental changes. Building from recent work showing that component steps in biogeochemical processes are encoded by separate microbial taxa [9, 10], we hypothesized that within-gene diversity varies between component steps, and further that temporal dynamics of diversity would vary between steps. Our metagenomic data from a dynamic groundwater-surface water mixing zone were consistent with this hypothesis and demonstrated that within-gene diversity, and the dynamics of that diversity, are variable across genes. This outcome suggests that a community’s taxonomic diversity or the abundance or diversity of any single (proxy) gene is not be a reliable predictor of stability in functional potential for multi-step biogeochemical processes, and that portions of the community that encode component steps with low within-gene diversity may be the most critical when considering potential decreases in function. Therefore, there is a need to shift the focus of analyses from taxonomic diversity or ‘representative’ gene abundances to a comprehensive understanding of within-gene diversity and dynamics across processes. Below we place these discoveries in context of previous work and point toward how they can be used to improve predictive models of system function.

Diversity dynamics of nitrification genes

The nitrification process showed low diversity for both steps examined, leading to the possibility that these activities are susceptible to loss or diminished function. Nitrification was originally described as a cooperative process, requiring an ammonia oxidizing organism that produces nitrite and a nitrite oxidizing organism that converts the nitrite to nitrate [27]. Recently, organisms have been identified that have both activities (comammox) [28]. The range of organisms known to encode nitrification activities is narrow, although it does include both Bacteria (Nitrosomonas and Nitrospira) and Archaea (Thaumarchaeota). The observed abundance of nitrifying organisms in sediment communities, both freshwater and marine, suggests nitrification is an important activity in the subsurface environment [18, 29, 30]. The limited taxonomic distribution of nitrification activities in the hyporheic community was expected, however the low diversity, one phylotype for *nxrA*, and one sequence apiece for the bacterial and archaeal *amoA*s is extreme. This lack of diversity suggests these activities could be unstable, given observations demonstrating that community-level functional stability increases with diversity [7, 31, 32]. However, we observed very stable abundance of these organisms across the seasonal shift in water chemistry, suggesting that the organisms encoding these activities are well adapted to the range of environmental conditions historically experienced by this community. Any extraordinary shift in biotic (*e.g.*, viruses, predation) or abiotic (*e.g.*, redox potential, temperature) conditions that selects against the small number of taxa involved in nitrification, however, could quickly degrade the community’s nitrification potential. With no other apparent organisms available to supplement or take over this role, this fundamental service could be degraded or lost from this community with unknown repercussions for the microbial community and the larger ecosystem [33, 34]. Recently, nitrifiers, and in particular Archaeal nitrifiers, have been shown to be active in carbon fixation in freshwater benthic sediments [35]. Thus, loss of nitrifiers could impact coupled carbon-nitrogen cycling in the subsurface and associated river corridors.

Diversity dynamics of denitrification genes

Denitrification genes have been identified in a broad range of taxa [36], and as such, our expectation was that within the hyporheic zone community there would be a high diversity across all component steps [37, 38]. While we did observe considerable overall abundance of all genes, the levels of richness for the genes representing the individual activities varied, ranging from 52 phylotypes for nitrate reduction (*narG*) to 23 phylotypes for nitrite reductase (*nirK* and *nirS*). This observation supports the concept that denitrification genes are distributed among members of the community as partial pathways or individual genes [14, 15]. Further, there was a surprising distribution of nitrous oxide reductase genes, with the type II form (*nosZII*), which is typically found in non-denitrifying organisms [25], having much greater abundance and richness (49 phylotypes) than the type I form (*nosZI*, 2 phylotypes).

Temporal variance of within-gene diversity for genes involved in both nitrification and denitrification demonstrates that the organisms encoding these activities are sensitive to different ecological selection pressures and thus different strategies are required to maintain functional potential in response to perturbation. For genes with high phylotype richness, high temporal abundance variance indicates a changing phylotype profile (*nirKS*, *norB*). These functions may be maintained through resilient microbial taxa that recover rapidly from environmental change. Conversely, low temporal variance (*narG*, *nosZ*) indicates a stable phylotype profile. These functions are maintained through resistant taxa that persist across a broad range of environmental conditions, with the possibility that the other low abundance phylotypes are capable of supplanting them should they fail under different conditions.

It is notable that while all genes associated with denitrification had high phylotype richness (in contrast to nitrification genes), the genes associated with intermediate reactions had higher temporal diversity variance than *narG* (Table 1), which encodes the initial step in denitrification (i.e., nitrate reduction). One explanation for the observed differences could be that there are different levels of competition for the substrates fueling each activity. Intermediate substrates nitrite and nitric oxide may be produced slowly and/or consumed quickly, especially considering there are multiple cellular processes for which they are intermediates and they are both toxic to cells. Supporting this contention, nitrite is typically undetectable in samples from this location, while nitrate is readily detectable [21]. Low availability would lead to high substrate competition, which could result in the increased phylotype turnover observed in *nirK*, *nirS* and *norB* genes. Modeling the redundancy provided to a process by within-gene diversity thus requires an understanding of temporal variation in the selective pressures for each gene involved.

Influence of seasonal changes in hydrogeochemistry

Seasonal changes in groundwater to surface water ratios appear to be a major influence on N-cycling functional potential in microbial communities. Increase in groundwater content corresponded to increasing per-capita abundance of nitrification genes and decreasing abundance/increasing diversity of denitrification genes. The *nirKS* and *norB* gene families, which displayed similar high phylotype turnover behavior, were not similar in their response to the environmental parameters measured, with *nirKS* showing a decrease in richness in response to groundwater while *norB* showed a decrease in abundance. The *narG* and *nosZ* gene families, which showed more stable profiles, both increased in diversity in response to groundwater, however, *nosZ* did so through increased richness, while *narG* likely gained evenness through reduced abundance of dominant phylotypes. Organic carbon (NPOC) had a much weaker association with gene-level metrics, relative to groundwater. A group of co-occurring organisms with a negative correlation to groundwater has been reported in this sediment system

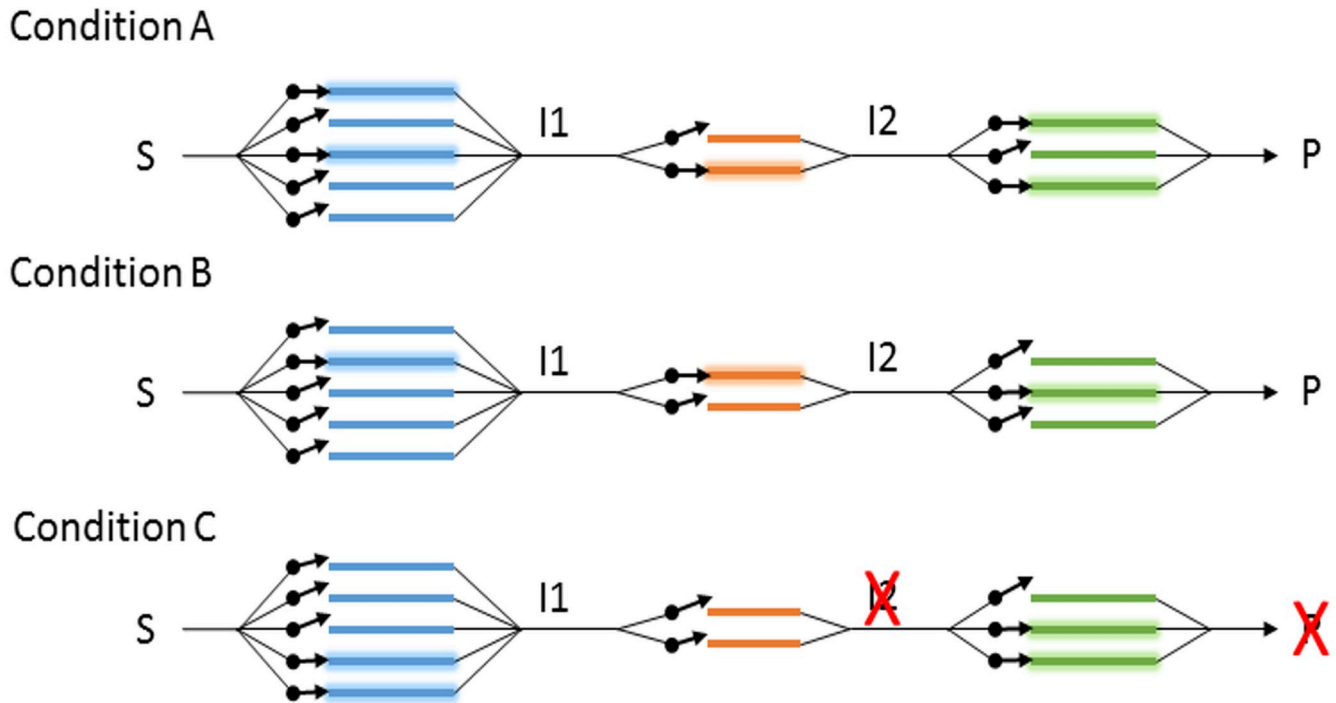


Fig 11. Circuit diagram of a metabolic pathway. Steps in series convert substrates (S), to various intermediates (I1, I2), to a product (P). Redundancy is represented by parallel paths, which can be regulated individually (denoted by arrow gates). Under conditions A and B, product is produced, but by different paths, whereas under condition C, although the blue and green steps are active, neither of the orange steps are, preventing production of I2 and P.

<https://doi.org/10.1371/journal.pone.0228165.g011>

[21]. The group is dominated by Alpha-, Beta- and Gammaproteobacteria, Bacteroidetes and Planctomycetes, the same taxa that encode nearly all of the identified denitrification genes. Strong homogenous selection was shown to be the mechanism structuring this group [20]. Taken together, these data suggest that some factor other than carbon that is within the groundwater is the selective force driving the diversity dynamics of these organisms carrying N-cycling genes. A likely candidate is the N content of groundwater, which is significantly higher than that of the surface water [26].

Gene diversity and process resilience

Conceptualizing and studying diversity within individual gene families is a departure from the contemporary perspective that largely focuses on organismal diversity or abundances of gene families. Variation in diversity across component steps of key biogeochemical processes and the dynamics of within-gene diversity in response to environmental change is therefore unexplored. This hampers our ability to predict ecosystem responses to future environmental changes. To illustrate the importance of diversity across individual component steps of biogeochemical processes, we use the analogy of an electrical circuit (Fig 11). Continuity from one step to the next is required for the full process/circuit to function. To preserve integrity of the circuit there is parallelization within each component step, whereby there are multiple options for completing a given step (Condition A). In a biological context, this manifests as multiple organisms encoding the same activity through different alleles of the same gene. Under different environmental conditions, various options may not be available either because the conditions are not favorable to the expression or operation of the gene, or the organism encoding that gene is eliminated from the community. The function is maintained by the availability or

introduction of alternates that can function under the new conditions (Condition B). Conditions may exist, however, under which no options for a given component step are available to the system, for example if an anaerobic system was exposed to sufficient oxygen to inhibit nitrous oxide reductase activity. This scenario will prevent the full biogeochemical process (e.g., denitrification) from completing, at least temporarily, even if some component steps are functioning (Condition C). Steps with low within-gene diversity are more likely to experience environmental conditions that cause all options to be eliminated. Just as a chain is only as strong as its weakest link, the ability of a metabolic pathway to continue functioning is determined by the component step with the lowest diversity.

We propose that accounting for the influence of environmental variation on realized biogeochemical rates in predictive models should connect environmental conditions to the dynamics of component steps. Doing so would allow models to account for variation in the susceptibility of each step to perturbation, based on within-gene diversity and dynamics. For example, reaction network models could represent the combined influence of gene-level abundance and diversity on continued function during and after perturbation. Recent modeling developments open up such opportunities, such as Song et al.'s reaction network model that explicitly represents control of enzyme expression at each step along a given biogeochemical pathway [39]. This model could be easily modified to represent different levels of diversity and abundance of gene phylotypes across component steps. Numerical experiments using the resulting model could comprehensively explore the sensitivity of biogeochemical function to among-step variation in within-gene diversity and dynamics. We also contend that there is a need to incorporate within-gene diversity into our conceptualization of diversity and focus on understanding the ecological processes governing diversity within individual genes. Merging such ecological knowledge with mechanistic biogeochemical models should improve our ability to predict biogeochemical function under future environmental conditions.

Experimental procedures

Sampling

Sediment communities were captured using sand packs incubated within piezometers as described [20]. Briefly, 1.2 m, fully-screened, stainless steel piezometers (5.25 cm inner diameter) (S5a Fig) were deployed along the margin of the Columbia River at approximately 46° 22' 15.80"N, 119° 16' 31.52"W. Sand packs composed of ~80 cm³ of locally-sourced medium grade sand (>0.425mm <1.7mm) packed into 2 x 4.5", 18/8 mesh stainless steel infuser plugged with Pyrex fiber glass (S5b Fig) were sterilized by combustion at 450°C for 8hr and then deployed in pairs for six week incubations collected at three week intervals from April 30, 2014 to November 25, 2014. Upon retrieval, paired sand packs were combined and homogenized. A ~145 mL subsample was flash-frozen and transported on dry ice back to the laboratory for metagenomic analysis. Aqueous samples were taken as previously described [20]. Briefly, at each piezometer, peristaltic pumps and manifolds were purged for 10–15 minutes. Following the purge, water was pumped through 0.22 μm polyethersulfone Sterivex filters for 30 minutes. Filtered water was used for water chemistry analysis.

Sampling equipment was installed after required consultations and permits were obtained from appropriate state and federal agencies, including the Department of Energy's Pacific Northwest Site Office, the U.S. Fish and Wildlife Service, the National Marine Fisheries Service, the U.S. Army Corps of Engineers, and the Washington Department of Fish and Wildlife. All federal requirements under the National Environmental Policy Act were followed.

Water chemistry

Water chemistry was determined as previously described [20]. Briefly, water temperature was measured with a handheld meter (Ultrameter II, Myron L Co Carlsbad, CA). A YSI Pro ODO handheld with an optical DO probe (YSI Inc. Yellow Springs, OH) was used to measure dissolved oxygen. NPOC was determined by the combustion catalytic oxidation/NDIR method using a Shimadzu TOC-Vcsh with ASI-V auto sampler (Shimadzu Scientific Instruments, Columbia, MD). Samples were acidified with 2 N HCl and sparged for 5 minutes to remove DIC. The sample was then injected into the furnace set to 680°C. Nitrate concentrations were determined on a Dionex ICS-2000 anion chromatograph with AS40 auto sampler. A 25-minute gradient method was used with a 25- μ L injection volume and a 1 mL/min flow rate at 30°C (EPA-NERL: 300.0).

DNA extraction

Genomic DNA was prepared from piezometer T4 sediment samples as previously described [20]. Briefly, to release biomass, thawed samples were suspended in 20mL of chilled PBS /0.1% Na-pyrophosphate solution and vortexed for 1 min. The suspended fraction was decanted to a fresh tube and centrifuged for 15' at 7000 x g at 10°C. DNA was extracted from the resulting pellets using the MoBio PowerSoil kit in plate format (MoBio Laboratories, Inc., Carlsbad, CA) following manufacturer's instructions with the addition of a 2-hour proteinase-K incubation at 55°C prior to bead-beating to facilitate cell lysis. Subsamples of each preparation were used for 16S rRNA amplicon sequencing and shotgun metagenomic sequencing.

Sequencing

Genomic DNA purified from sandpack samples was submitted to the Joint Genome Institute under JGI/EMSL proposal 1781 for paired-end sequencing on an Illumina HiSeq 2500 sequencer. Results from the sequencing are presented in S1 Table. Data sets are available through the JGI Genome Portal (<http://genome.jgi.doe.gov>). Project identifiers are listed in S1 Table.

For the 16S rRNA amplicon analysis, the protocol developed by the Earth Microbiome Project (<http://press.igsb.anl.gov/earthmicrobiome/emp-standard-protocols/16s/>) was followed, with the exception that the twelve base barcode sequence was included in the forward primer. Amplicons were sequenced on an Illumina MiSeq using the 300 cycle MiSeq Reagent Kit v2 (<http://www.illumina.com/>) according to manufacturer's instructions.

Metagenomic analysis

To quantitate gene families of interest, hidden Markov models (HMMs) were obtained or built and searched against raw metagenomic reads. HMMs used in this study are listed in Table 2. HMMs were searched against raw reads using MaxRebo (Lee Ann McCue, unpubl.), which translates each read in six frames, and searches the translations against the target HMM(s), using HMMer [40] on a distributed, high-performance computing framework. Output was screened for reads with a significant score ($e\text{-value} \leq 1e\text{-}25$) against the HMM. Raw counts were converted to RPKM (reads per kilobase of gene length per million reads) using the HMM length x 3 as the gene length. Results from forward and reverse reads were averaged and normalized against the summed RPKMs of the rplB and rplB_arch models. Individual genes of interest were assembled from the combined metagenomic datasets using the Xander assembler [41] and the HMMs listed in Table 2 and associated required files. Resulting contigs were clustered at 90% amino acid identity (Supplementary Data 1) to define phylotypes. Phylogeny was

Table 2. HMMs used in this study.

Gene	HMM	Source
Nitrate reductase, alpha subunit (narG)	narG	FunGene ¹
Cu-containing nitrite reductase (nirK)		
Clade I	nirK1	PNNL ²
Clade II	nirK2	PNNL
Fe-containing nitrite reductase (nirS)	nirS	FunGene
Nitric oxide reductase (norB)		
Copper	norB_cNor	FunGene
quinone	norB_qNor	FunGene
Nitrous oxide reductase		
nosZI	nosZ	FunGene
'non-denitrifying' (nosZII)	nosZ_a2	FunGene
Ammonia monooxygenase		
bacterial	amoA_AOB	FunGene
archaeal	amoA_AOA	FunGene
Nitrite oxidoreductase, alpha subunit	nxrA-1	PNNL
Ribosomal protein RplB		
bacterial	rplB	FunGene
archaeal	rplB_arch	PNNL

¹ Available at https://github.com/rdpstaff/Xander_assembler

² Available at https://github.com/wichne/Xander_files

<https://doi.org/10.1371/journal.pone.0228165.t002>

assessed by aligning protein sequences with mafft v7.164b [42, 43] and constructing approximated maximum-likelihood trees using FastTree v2.1.9 [44]. Phylotype abundance profiles were determined by searching individual metagenomic read sets against the resulting gene contigs and calculating RPKM values and normalizing against the summed phylotype RPKM for the gene. Bray-Curtis dissimilarity between samples for each gene was calculated using the R package vegan [45], and resulting values were used to generate a boxplot.

Community analysis

Amplicon data used was from Graham et al., 2016b. Bray-Curtis distance was determined as described below, and plotted using R.

Statistics

Bray-Curtis dissimilarity, as implemented in the R package vegan [45], was used to measure beta diversity. Values were averaged for both the total dataset and the T4 dataset alone. Early (n = 6) versus late (n = 5) gene abundance comparisons were tested for significance using the Mann-Whitney-Wilcoxon test as implemented in R v.3.3.2 (<https://www.r-project.org>). For turnover heatmaps, assembled sequences were searched against the read set to estimate individual abundances. Sequences were then clustered into phylotypes at 90% identity, and abundances summed. The relative abundance of each phylotype was then determined by dividing its abundance by the summed abundance of all phylotypes of the gene in question. Trees were determined from nucleic acid sequence alignments (mafft v) using the maximum-likelihood approach implemented in FastTree. Inverse Simpson statistic for the assembled sequences was calculated cumulatively for each gene at each time point, also using the vegan package. Linear regressions and associated R² and p-values were calculated in R v3.3.2.

Supporting information

S1 Fig. Environmental parameter correlation. Temperature (Temp), dissolved oxygen (DO), chloride ion concentration (Cl), sulfate concentration (SO₄), nitrate concentration (NO₃) and dissolved organic carbon (measured as non-purgeable organic carbon, NPOC) measurements were taken for all samples. Pair-wise correlation of observations were performed to determine the independence of the parameters.

(PDF)

S2 Fig. Environmental parameter vs diversity (as measured by the inverse Simpson statistic) linear regression analysis. Blue borders: $p < 0.10$; Orange borders: $p < 0.05$.

(PDF)

S3 Fig. Environmental parameter vs gene abundance linear regression analysis. Blue borders: $p < 0.10$; Orange borders: $p < 0.05$.

(PDF)

S4 Fig. Environmental parameter vs richness linear regression analysis. Blue borders: $p < 0.10$; Orange borders: $p < 0.05$.

(PDF)

S5 Fig. Sampling setup. (a) Stainless steel piezometers (5.25 cm inner diameter) that were fully-screened for 1.2 m were driven into the river bottom sediment. (b) 4.5" stainless steel infusers (18/8 mesh) were packed with ~80 cm³ of locally-sourced medium grade sand (>0.425mm <1.7mm) and plugged with Pyrex fiber glass. Paired sand packs were deployed as shown in panel a) for six week incubations collected at three week intervals from April 30, 2014 to November 25, 2014.

(PDF)

S1 Table. Metagenomic sequence data sets.

(PDF)

Acknowledgments

A portion of the research was performed using Institutional Computing at PNNL.

Author Contributions

Conceptualization: William C. Nelson, Emily B. Graham, James C. Stegen.

Data curation: William C. Nelson.

Formal analysis: William C. Nelson.

Funding acquisition: James C. Stegen.

Investigation: William C. Nelson, Emily B. Graham, Alex R. Crump, Sarah J. Fansler, Evan V. Arntzen, David W. Kennedy.

Methodology: William C. Nelson.

Project administration: James C. Stegen.

Supervision: James C. Stegen.

Writing – original draft: William C. Nelson, Emily B. Graham, James C. Stegen.

Writing – review & editing: William C. Nelson, Emily B. Graham, James C. Stegen.

References

1. Gibbons SM, Gilbert JA. Microbial diversity—exploration of natural ecosystems and microbiomes. *Curr Opin Genet Dev*. 2015; 35:66–72. Epub 2015/11/26. <https://doi.org/10.1016/j.gde.2015.10.003> PMID: 26598941.
2. Walker BH. Biodiversity and Ecological Redundancy. *Conserv Biol*. 1992; 6(1):18–23. <https://doi.org/10.1046/j.1523-1739.1992.610018.x>
3. Yachi S, Loreau M. Biodiversity and ecosystem productivity in a fluctuating environment: The insurance hypothesis. *Proc Natl Acad Sci USA*. 1999; 96(4):1463–8. <https://doi.org/10.1073/pnas.96.4.1463> PMID: 9990046
4. Torsvik V, Ovreas L. Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol*. 2002; 5(3):240–5. Epub 2002/06/12. [https://doi.org/10.1016/s1369-5274\(02\)00324-7](https://doi.org/10.1016/s1369-5274(02)00324-7) PMID: 12057676.
5. Rosenfeld JS. Functional redundancy in ecology and conservation. *Oikos*. 2002; 98(1):156–62. <https://doi.org/10.1034/j.1600-0706.2002.980116.x>
6. Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, et al. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecol Monogr*. 2005; 75(1):3–35. <https://doi.org/10.1890/04-0922>
7. Allison SD, Martiny JB. Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A*. 2008; 105 Suppl 1:11512–9. PMID: 18695234.
8. Shade A, Peter H, Allison SD, Baho DL, Berga M, Burgmann H, et al. Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*. 2012; 3. ARTN 417 <https://doi.org/10.3389/fmicb.2012.00417> PMID: 23267351
9. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun*. 2016; 7:13219. <https://doi.org/10.1038/ncomms13219> PMID: 27774985.
10. Mobberley JM, Lindemann SR, Bernstein HC, Moran JJ, Renslow RS, Babauta J, et al. Organismal and spatial partitioning of energy and macronutrient transformations within a hypersaline mat. *Fems Microbiol Ecol*. 2017; 93(4). Epub 2017/03/24. <https://doi.org/10.1093/femsec/fix028> PMID: 28334407.
11. West SA, Cooper GA. Division of labour in microorganisms: an evolutionary perspective. *Nature Reviews Microbiology*. 2016; 14:716. <https://doi.org/10.1038/nrmicro.2016.111> PMID: 27640757
12. Zumft WG. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev*. 1997; 61(4):533–616. Epub 1997/12/31. PMID: 9409151.
13. Graf DRH, Jones CM, Hallin S. Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community Structure for N₂O Emissions. *Plos One*. 2014; 9(12). <https://doi.org/10.1371/journal.pone.0114118> PMID: 25436772
14. Bru D, Ramette A, Saby NP, Dequiedt S, Ranjard L, Jolivet C, et al. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J*. 2011; 5(3):532–42. Epub 2010/08/13. <https://doi.org/10.1038/ismej.2010.130> PMID: 20703315.
15. Keil D, Meyer A, Berner D, Poll C, Schutzenmeister A, Piepho HP, et al. Influence of land-use intensity on the spatial distribution of N-cycling microorganisms in grassland soils. *Fems Microbiol Ecol*. 2011; 77(1):95–106. Epub 2011/03/18. <https://doi.org/10.1111/j.1574-6941.2011.01091.x> PMID: 21410493.
16. Nelson MB, Berlemont R, Martiny AC, Martiny JB. Nitrogen Cycling Potential of a Grassland Litter Microbial Community. *Appl Environ Microbiol*. 2015; 81(20):7012–22. <https://doi.org/10.1128/AEM.02222-15> PMID: 26231641.
17. Nelson MB, Martiny AC, Martiny JB. Global biogeography of microbial nitrogen-cycling traits in soil. *Proc Natl Acad Sci U S A*. 2016; 113(29):8033–40. <https://doi.org/10.1073/pnas.1601070113> PMID: 27432978.
18. Stoliker DL, Repert DA, Smith RL, Song BK, LeBlanc DR, McCobb TD, et al. Hydrologic Controls on Nitrogen Cycling Processes and Functional Gene Abundance in Sediments of a Groundwater Flow-Through Lake. *Environ Sci Technol*. 2016; 50(7):3649–57. <https://doi.org/10.1021/acs.est.5b06155> PMID: 26967929
19. Graham EB, Wieder WR, Leff JW, Weintraub SR, Townsend AR, Cleveland CC, et al. Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biol Biochem*. 2014; 68:279–82. <https://doi.org/10.1016/j.soilbio.2013.08.023>
20. Graham EB, Crump AR, Resch CT, Fansler S, Arntzen E, Kennedy DW, et al. Coupling Spatiotemporal Community Assembly Processes to Changes in Microbial Metabolism. *Frontiers in Microbiology*. 2016; 7(1949). <https://doi.org/10.3389/fmicb.2016.01949> PMID: 28123379

21. Graham EB, Crump AR, Resch CT, Fansler S, Arntzen E, Kennedy DW, et al. Deterministic influences exceed dispersal effects on hydrologically-connected microbiomes. *Environ Microbiol.* 2017; 19(4):1552–67. <https://doi.org/10.1111/1462-2920.13720> PMID: 28276134.
22. Graham EB, Crump AR, Resch CT, Fansler S, Arntzen E, Kennedy DW, et al. Coupling Spatiotemporal Community Assembly Processes to Changes in Microbial Metabolism. *Front Microbiol.* 2016. <https://doi.org/10.3389/fmicb.2016.01949> PMID: 28123379
23. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A.* 2005; 102(7):2567–72. Epub 2005/02/11. <https://doi.org/10.1073/pnas.0409727102> PMID: 15701695.
24. Jones CM, Spor A, Brennan FP, Breuil MC, Bru D, Lemanceau P, et al. Recently identified microbial guild mediates soil N₂O sink capacity. *Nat Clim Change.* 2014; 4(9):801–5. <https://doi.org/10.1038/Nclimate2301>
25. Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-Garcia C, et al. Unexpected non-denitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc Natl Acad Sci U S A.* 2012; 109(48):19709–14. <https://doi.org/10.1073/pnas.1211238109> PMID: 23150571.
26. Stegen JC, Johnson T, Fredrickson JK, Wilkins MJ, Konopka AE, Nelson WC, et al. Influences of organic carbon speciation on hyporheic corridor biogeochemistry and microbial ecology. *Nat Commun.* 2018; 9(1):585. Epub 2018/02/10. <https://doi.org/10.1038/s41467-018-02922-9> PMID: 29422537.
27. Winogradsky S. Recherches sur les organismes de la nitrification. *Ann Inst Pasteur.* 1890; 4(4):213–31.
28. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by *Nitrospira* bacteria. *Nature.* 2015; 528(7583):504–9. <https://doi.org/10.1038/nature16461> PMID: 26610024.
29. Wang ZY, Qi Y, Wang J, Pei YS. Characteristics of aerobic and anaerobic ammonium-oxidizing bacteria in the hyporheic zone of a contaminated river. *World J Microb Biot.* 2012; 28(9):2801–11. <https://doi.org/10.1007/s11274-012-1090-6> PMID: 22806720
30. Lansdown K, Heppell CM, Dossena M, Ullah S, Heathwaite AL, Binley A, et al. Fine-Scale in Situ Measurement of Riverbed Nitrate Production and Consumption in an Armored Permeable Riverbed. *Environ Sci Technol.* 2014; 48(8):4425–34. <https://doi.org/10.1021/es4056005> PMID: 24628544
31. Girvan MS, Campbell CD, Killham K, Prosser JI, Glover LA. Bacterial diversity promotes community stability and functional resilience after perturbation. *Environ Microbiol.* 2005; 7(3):301–13. <https://doi.org/10.1111/j.1462-2920.2005.00695.x> PMID: 15683391
32. Tilman D, Knops J, Wedin D, Reich P, Ritchie M, Siemann E. The influence of functional diversity and composition on ecosystem processes. *Science.* 1997; 277(5330):1300–2. <https://doi.org/10.1126/science.277.5330.1300>
33. Dobson A, Lodge D, Alder J, Cumming GS, Keymer J, McGlade J, et al. Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology.* 2006; 87(8):1915–24. Epub 2006/08/30. [https://doi.org/10.1890/0012-9658\(2006\)87\[1915:hltcat\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1915:hltcat]2.0.co;2) PMID: 16937628.
34. Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, et al. Impacts of biodiversity loss on ocean ecosystem services. *Science.* 2006; 314(5800):787–90. Epub 2006/11/04. <https://doi.org/10.1126/science.1132294> PMID: 17082450.
35. Coskun OK, Pichler M, Vargas S, Gilder S, Orsi WD. Linking Uncultivated Microbial Populations and Benthic Carbon Turnover by Using Quantitative Stable Isotope Probing. *Appl Environ Microbiol.* 2018; 84(18). Epub 2018/07/08. <https://doi.org/10.1128/AEM.01083-18> PMID: 29980553.
36. Shapleigh JP. Denitrifying Prokaryotes. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. *The Prokaryotes: Prokaryotic Physiology and Biochemistry.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. p. 405–25.
37. Graham EB, Knelman JE, Schindlbacher A, Siciliano S, Breulmann M, Yannarell A, et al. Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? *Front Microbiol.* 2016; 7:214. <https://doi.org/10.3389/fmicb.2016.00214> PMID: 26941732.
38. Schimel J. Ecosystem Consequences of Microbial Diversity and Community Structure. *Ecol Stu An.* 1995; 113:239–54.
39. Song HS, Goldberg N, Mahajan A, Ramkrishna D. Sequential computation of elementary modes and minimal cut sets in genome-scale metabolic networks using alternate integer linear programming. *Bioinformatics.* 2017; 33(15):2345–53. <https://doi.org/10.1093/bioinformatics/btx171> PMID: 28369193
40. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol.* 2011; 7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195> PMID: 22039361.
41. Wang Q, Fish JA, Gilman M, Sun Y, Brown CT, Tiedje JM, et al. Xander: employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome.* 2015; 3:32. <https://doi.org/10.1186/s40168-015-0093-6> PMID: 26246894.

42. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002; 30(14):3059–66. <https://doi.org/10.1093/nar/gkf436> PMID: 12136088.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690.
44. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 2010; 5(3):e9490. <https://doi.org/10.1371/journal.pone.0009490> PMID: 20224823.
45. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003; 14(6):927–30. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>