



Inferring the Contribution of Microbial Taxa and Organic Matter Molecular Formulas to Ecological Assembly

Robert E. Danczak^{1*}, Aditi Sengupta², Sarah J. Fansler¹, Rosalie K. Chu³, Vanessa A. Garayburu-Caruso¹, Lupita Renteria¹, Jason Toyoda³, Jacqueline Wells¹ and James C. Stegen¹

¹Ecosystem Sciences, Pacific Northwest National Laboratory, Richland, WA, United States, ²Department of Biology, California Lutheran University, Thousand Oaks, CA, United States, ³Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, United States

OPEN ACCESS

Edited by:

Daliang Ning,
University of Oklahoma,
United States

Reviewed by:

Fangqiong Ling,
Washington University in St. Louis,
United States
Oskar Modin,
Chalmers University of Technology,
Sweden

*Correspondence:

Robert E. Danczak
robert.danczak@pnl.gov

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 28 October 2021

Accepted: 14 January 2022

Published: 18 February 2022

Citation:

Danczak RE, Sengupta A, Fansler SJ, Chu RK, Garayburu-Caruso VA, Renteria L, Toyoda J, Wells J and Stegen JC (2022) Inferring the Contribution of Microbial Taxa and Organic Matter Molecular Formulas to Ecological Assembly. *Front. Microbiol.* 13:803420. doi: 10.3389/fmicb.2022.803420

Understanding the mechanisms underlying the assembly of communities has long been the goal of many ecological studies. While several studies have evaluated community wide ecological assembly, fewer have focused on investigating the impacts of individual members within a community or assemblage on ecological assembly. Here, we adapted a previous null model β -nearest taxon index (β NTI) to measure the contribution of individual features within an ecological community to overall assembly. This new metric, called feature-level β NTI (β NTI_{feat}), enables researchers to determine whether ecological features (e.g., individual microbial taxa) contribute to divergence, convergence, or have insignificant impacts across spatiotemporally resolved metacommunities or meta-assemblages. Using β NTI_{feat}, we revealed that unclassified microbial lineages often contributed to community divergence while diverse groups (e.g., Crenarchaeota, Alphaproteobacteria, and Gammaproteobacteria) contributed to convergence. We also demonstrate that β NTI_{feat} can be extended to other ecological assemblages such as organic molecules comprising organic matter (OM) pools. OM had more inconsistent trends compared to the microbial community though CHO-containing molecular formulas often contributed to convergence, while nitrogen and phosphorus-containing formulas contributed to both convergence and divergence. A network analysis was used to relate β NTI_{feat} values from the putatively active microbial community and the OM assemblage and examine potentially common contributions to ecological assembly across different communities/assemblages. This analysis revealed that P-containing formulas often contributed to convergence/divergence separately from other ecological features and N-containing formulas often contributed to assembly in coordination with microorganisms. Additionally, members of Family *Geobacteraceae* were often observed to contribute to convergence/divergence in conjunction with both N- and P-containing formulas, suggesting a coordinated ecological role for family members and the nitrogen/phosphorus cycle. Overall, we show that β NTI_{feat} offers opportunities to investigate the community or

assemblage members, which shape the phylogenetic or functional landscape, and demonstrate the potential to evaluate potential points of coordination across various community types.

Keywords: community assembly, β -nearest taxon index, null modeling, FTICR-MS, metacommunity ecology, meta-metabolome ecology

INTRODUCTION

Evaluating the processes which govern community diversity is often the goal of ecological studies across all ecosystems (Swenson et al., 2006; Kraft et al., 2007; Gilbert and Bennett, 2010; Smith and Lundholm, 2010; Chase and Myers, 2011; George et al., 2011; Stegen et al., 2013; Herren and McMahon, 2017; Zhou and Ning, 2017; Danczak et al., 2020b). Analogously, researchers have also focused on understanding the processes governing the composition of organic molecules or metabolites within organic matter (OM) assemblages (Danczak et al., 2020a, 2021). While methods might vary in how researchers investigate these processes (e.g., variation partitioning, trait-based analyses, and null modeling), each study attempts to determine when, where, and how various ecological assembly processes give rise to specific community/assemblage configurations. By better understanding the distribution of these processes and the circumstances under which they dominate, we will be able to better understand the fundamental principles governing community/assemblage structure. However, less attention has been paid to the impact of ecological processes on individual community/assemblage members or to the impact of individual members on ecological assembly. Hereafter, in order to limit confusion, both biological and chemical members are referred to as “features,” while biological communities and OM assemblages are referred to as “communities” (Table 1).

Evaluating the impacts of individual features provides several benefits to ecological researchers. Firstly, feature-level metrics provide researchers the opportunity to investigate how specific

features (e.g., a microbial taxon or metabolite) contribute to assembly under varied environmental regimes or across different spatiotemporal scales. As detailed by Ning et al. (2020), members of a given taxonomic level can respond differently to environmental stresses, with some experiencing variable selection and others being limited by dispersal. Feature-level metrics will allow researchers to disentangle the contributions (or lack thereof) of individual members and identify groups putatively most relevant to community assembly. The ability to observe individual contributions would allow researchers to evaluate the ecological roles of specific organisms and/or organic molecules in the absence of explicit physiological or biochemical information based on traits inferred from phylogeny/taxonomy. For example, Ning et al. (2020) observed that a group of Bacillales significantly experienced homogeneous selection in hot/dry environments potentially due to their enhanced survivability.

Secondly, feature-level metrics will allow ecologists to compare and relate the contributions to assembly dynamics across community types in order to observe potential ecological coordination. For example, these metrics can be directly compared across different community types to find groups of cross-community features, which exert coordinated control on the larger metacommunity/meta-assemblage, potentially highlighting a common ecological pressure or ecological interaction (e.g., a specific environmental condition driving the selection of features across community boundaries). Investigating how assembly compares across disparate groups has enabled a deeper understanding of the fundamental factors structuring ecological communities. Danczak et al. (2020b) revealed that viral and microbial communities experienced coordinated assembly processes despite facing separate environmental pressures in a fractured shale ecosystem. Jiao et al. (2020) demonstrated that the balance of cross-kingdom species interactions across Archaea-Bacteria-Fungi mediated community assembly in an agricultural soil ecosystem. These examples indicate that the ability to measure and relate feature-level contributions to ecological assembly will help identify components disproportionately impacting phylogenetic or functional community structure.

We propose that a new metric called feature-level β -nearest taxon index ($\beta\text{NTI}_{\text{feat}}$) based upon an existing null modeling framework (βNTI) will provide these benefits. βNTI is particularly capable in assessing the assembly dynamics associated with ecological metacommunities and OM meta-metabolomes/assemblages and we show that it can be adapted to feature-level analyses. βNTI has been used extensively to study assembly processes. For example, researchers have revealed relationships between microbial community development and organic matter degradation (Stegen et al., 2016, 2018), coordination of assembly

TABLE 1 | Table of terms used throughout the manuscript and their definitions.

Terms	Definitions
Feature	<i>Feature</i> is used liberally in this manuscript. This term is meant to describe anything that can be described in a relational dendrogram (e.g., microbial community members and FTICR-MS molecular formula).
Community	<i>Community</i> refers to a collection of ecological features. Here, that describes both microbial communities and organic matter assemblages.
Contribution	$ \beta\text{NTI}_{\text{feat}} > 1$, with significant contributions inferred when $ \beta\text{NTI}_{\text{feat}} > 2$; When a feature “contributes” to ecological assembly, it exerts some deterministic impact on the community.
Contribution to convergence	$\beta\text{NTI}_{\text{feat}} < -2$; Features that “contribute to convergence” are those which drive ecological/functional similarities across communities.
Contribution to divergence	$\beta\text{NTI}_{\text{feat}} > 2$; Features that “contribute to divergence” are those which drive ecological/functional differences across communities.

between viral and microbial communities (Danczak et al., 2020b), the balance of niche- and dispersal-based processes the soybean microbiome (Moroenyane et al., 2021), and emphasized the importance of salinity in the assembly of desert microbial communities (Zhang et al., 2019).

Recently, Ning et al. (2020) developed an iterative null model based on β -net relatedness index (β NRI) called iCAMP, which represents a potential route to identify these feature-level dynamics. By first identifying the minimum phylogenetic level at which a phylogenetic signal exists (e.g., a relationship between evolutionary history and niche occupancy; Blomberg et al., 2003; Stegen et al., 2012), iCAMP groups community members and measures the ecological pressures acting upon that level. Using these metrics calculated across the phylogenetic tree, iCAMP can then estimate the balance of assembly processes acting on the community. This method is an excellent way to account for phylogenetic groups experiencing varied assembly processes in whole community analyses and represents a novel way to follow sub-community assembly through time or space. However, while capable of identifying assembly processes at levels below the entire community, this approach still investigates processes impacting assembly at the subcommunity level rather than measuring the degree which an individual feature impacts or is impacted by assembly. Fodelianakis et al. (2021) instead took an approach, called “phyloscore analysis” that focuses instead on the ecological contributions of specific taxa within a microbial community. Likewise, β NTI_{feat} focuses on individual features to measure their ecological contribution to community dynamics and highlights a point of complementarity across community, subcommunity, and feature-level foci (Table 1).

Feature-level β -nearest taxon index provides insight into the degree to which each observed feature contributes to either ecological convergence or divergence (Table 1). Ecological convergence occurs when some feature drives similarities in the phylogenetic or functional landscape across samples or within a dataset. Such a feature would be more phylogenetically or functionally conserved than expected by random chance. In contrast, ecological divergence occurs when a feature drives phylogenetic or functional differences across samples or a dataset; these features are more divergent than expected by random chance. Based on these interpretations, β NTI_{feat} stands as the phylogenetic or dendrogram-based complement to taxonomic metrics like SIMPER (Clarke, 1993). Importantly, given that β NTI_{feat} does not rely on abundance or taxonomic-based distance metrics, it is able to overcome many of the limitations associated with SIMPER (Warton et al., 2012). When compared to the phyloscore metric described by Fodelianakis et al. (2021), it is much more similar though differs in some mathematical specifics (namely the null implementation) and in its application across different scales.

Here, we describe the theory behind the β NTI_{feat} calculations, apply it to microbial and environmental metabolomic data, and discuss how interpretations vary with scale and dataset. First, we reveal that β NTI_{feat} can identify microbial taxa (down to the amplicon sequence variant, or ASV, level) and environmental metabolites (down to the specific molecular formulas), which disproportionately contribute to the ecological

structure of the respective community. Second, we demonstrate that we can track features with disproportionate contributions through time and relate them to each other to uncover ASVs and molecular formulas that have coordinated contributions to the biological and chemical composition of the study system.

MATERIALS AND METHODS

Sample Collection

Detailed sample collection is outlined in Sengupta et al. (2021a) but will be described briefly here. Sediments were collected from the hyporheic zone of the Columbia River shoreline in eastern Washington state on 14 January 2019 at 9 a.m. Pacific Standard Time. Five samples were collected from within a meter range, combined to make a composite sediment sample, and then sieved on site through a 2 mm sieve into a glass beaker. Sieved sediment was kept on blue ice for 30 min until transported back to the lab, where it was stored at 4°C until experimentation.

Experimental Design

A detailed experimental design is outlined in Sengupta et al. (2021a) but will be described briefly here. Sieved sediment was partitioned into two sets of vials, each vial containing 10 g of sediment: one set of vials were under inundated conditions for 23 days, and the other set were allowed to dry for 23 days. Once the vials were permitted to acclimate, they were subject treatment regimes designed around a series of wet/dry transition periods. While the details are important, these regimes largely translate to two bulk treatments based upon the number of days left dry: cumulatively dry (34, 31, and 27 days) and cumulatively inundated (0, 4, and 8 days). This resulted in a total of 20 vials per cumulative treatment.

16S rRNA Gene Sequencing and Processing

Detailed DNA and cDNA extraction methods can be found in Sengupta et al. (2021a). Briefly, incubated sediments were centrifuged and flash-frozen. gDNA was extracted following the protocol established by Bottos et al. (2018) RNA was extracted using the Qiagen PowerSoil RNA extraction kit (Qiagen, Germantown, MD), treated for contaminate DNA using DNase, quantified using a Qubit RNA kit (Thermo Fisher, Waltham, MA), and reverse transcribed into cDNA using the SuperScript™ IV First-Strand Synthesis System (Thermo Fisher Scientific, Waltham, MA). Amplicon sequencing for both gDNA and cDNA was performed following The Earth Microbiome Project. Sequences are accessible at NCBI's Sequence Read Archive using the accession number # PRJNA641165.

Sequence processing was performed using QIIME2 (Bolyen et al., 2019). Raw amplicon sequences were imported into the QIIME2 environment, denoised using DADA2 (*q2-dada2*), and assigned taxonomy using the SILVA v138 database (*q2-feature-classifier*; Quast et al., 2013; Callahan et al., 2016; Bokulich et al., 2018). A phylogenetic tree was generated by first aligning

amplicons using the MAFFT aligner (Kato, 2002) and then generating a maximum-likelihood tree (*q2-phylogeny*). The 16S rRNA gene amplicon maximum-likelihood tree is stored as **Supplementary File 1**.

FTICR-MS Analysis

Detailed Fourier transform ion cyclotron resonance mass spectrometer (FTICR-MS) data acquisition is outlined in Danczak et al. (2020a) and Sengupta et al. (2021a) but will be described briefly here. In brief, a solid-phase extraction on a PPL cartridge (Bond Elut) was used to concentrate carbon and remove salt (Dittmar et al., 2008). Extracted samples were injected into a 12 Tesla (12T) Bruker Solarix FTICR-MS outfitted with a standard electrospray ionization source (ESI) configured to negative mode. Resulting spectra were processed using the Bruker Daltonik Data Analysis software (v4.2) to obtain a peak list; Formularity was then used to assign molecular formulas to detected peaks following the Compound Identification Algorithm (Kujawinski and Behn, 2006; Tolić et al., 2017). The report generated in Formularity was then processed using the *ftmsRanalysis* R package to calculate various molecular properties (e.g., double-bond equivalents, modified aromaticity index, nominal oxidation state of carbon, Kendrick's defect, etc.) and assign compound classes (Hughey et al., 2001; Kim et al., 2003; Koch and Dittmar, 2006; LaRowe and Van Cappellen, 2011; Bailey et al., 2017; Rivas-Ubach et al., 2018; Bramer et al., 2020). Using the methods outlined in Danczak et al. (2020a), we generated a molecular characteristics dendrogram (MCD) for use in dendrogram-informed analyses; the MCD is stored as **Supplementary File 2**.

Ecological Analyses

Multivariate differences across the microbial community and OM assemblage were detected using ordinations in combination with PERMANOVA statistics (*adonis*; *vegan* package v2.5-7; Oksanen et al., 2019). A principal coordinate analysis (PCoA; *pcoa*; *ape* package v5.5) was generated using both Bray–Curtis dissimilarity (*vegdist*; *vegan* package v2.5-7) and β -mean nearest taxon distance (β MNTD; *comdistnt*; *picante* package v1.8.2) were calculated for both the microbial community and OM assemblage (Kembel et al., 2010; Oksanen et al., 2019; Paradis and Schliep, 2019).

The β NTI_{feat} Calculation

The β NTI_{feat} is intrinsically linked to the β NTI calculation and helps us understand the relationship between observed dendrogram-based relationships of individual features and some null expectation (**Figure 1**). First, β MNTD_{feat}, the minimum relational distance of a feature in one community to the nearest feature in another, needs to be calculated for the observed community across the entire dataset:

$$\beta\text{MNTD}_{\text{feat}} = \frac{1}{n} \sum_{j=1}^n f_{a_i} \min(d_{a,b_j}) \quad (1)$$

where f_{a_i} is the relative abundance of feature a in community i , n is the number of communities/samples in the dataset, and $\min(d_{a,b_j})$ is the average minimum relational distance

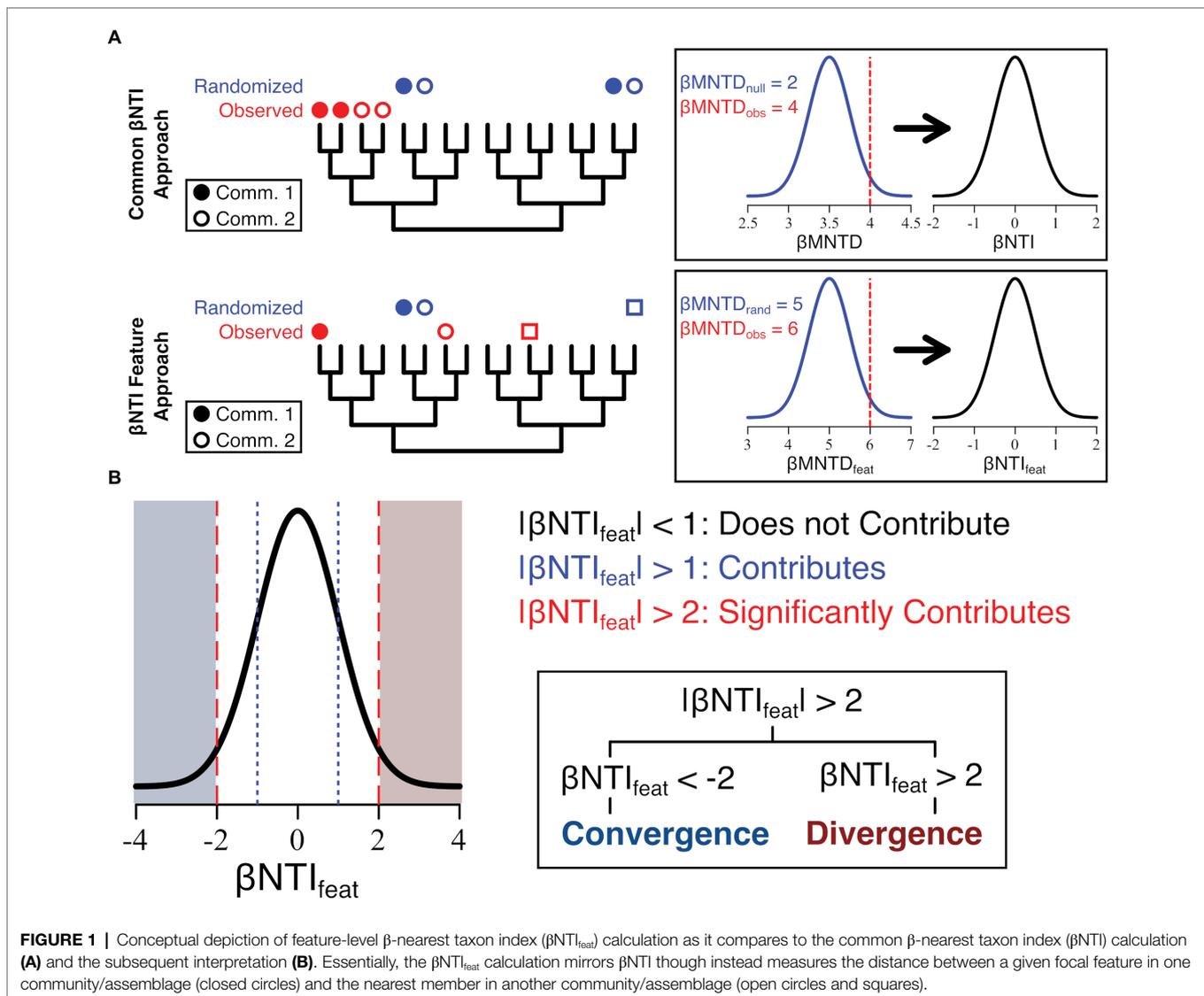
(e.g., the distance between tips on a dendrogram—equivalent to phylogenetic distance) of the fixed feature a in the fixed community i to any feature b in all communities j . In these current analyses, we are allowing conspecifics in our calculations, but they can be excluded pending experimental design. In this case, a conspecific feature is one which is present across both halves of a pairwise comparison and the inclusion/exclusion implementation matches that of the *comdistnt* in the *picante* R package (Kembel et al., 2010). In practice, this metric measures the average minimum distance between a given feature in one community and all other features in other communities. The key departure from the standard β MNTD calculation is that this calculation occurs from a fixed perspective; only one community is compared to all other communities at a single time. As with the traditional β NTI calculation, β MNTD_{feat} was also calculated for 999 randomized communities, which were generated by shuffling the tips of the provided dendrogram/phylogenetic tree using the function *taxaShuffle* from the *picante* R package (Kembel et al., 2010). Additionally, the null β MNTD_{feat} calculation has small amounts of phylogenetic noise (e.g., 1×10^{-20} – 5×10^{-20}) injected into them to allow for features present across both halves of the pairwise comparison to be included. By combining the null results with our observed results, we can calculate β NTI_{feat}:

$$\beta\text{NTI}_{\text{feat}} = \frac{\beta\text{MNTD}_{\text{feat}}^{\text{obs}} - \overline{\beta\text{MNTD}_{\text{feat}}^{\text{null}}}}{\beta\text{MNTD}_{\text{feat}}^{\text{sd}}} \quad (2)$$

where $\beta\text{MNTD}_{\text{feat}}^{\text{obs}}$ is the observed β MNTD_{feat} measurement, $\overline{\beta\text{MNTD}_{\text{feat}}^{\text{null}}}$ is the average β MNTD_{feat} for the null results, and $\beta\text{MNTD}_{\text{feat}}^{\text{sd}}$ is the SD of $\beta\text{MNTD}_{\text{feat}}^{\text{null}}$ values.

Estimating Feature-Level Ecological Dynamics

Following the same philosophy underlining the typical β NTI interpretations, β NTI_{feat} seeks to quantify the ecological processes occurring within and across communities. When describing the underlying theory of β NTI_{feat}, we use the term “community” as the general term to describe any assemblage of ecological data (**Table 1**). However, we want to stress that this method can be used on any relational set of data (e.g., microbial communities and DOM assemblages). Unlike β NTI, which evaluates whole communities through space or time, β NTI_{feat} is focused on identifying the contributions to community assembly by individual community members (**Figure 1**). For example, while β NTI is well-suited to identify variable selection as a process driving differences between two communities, it cannot measure which community members may be driving that variable selection. In contrast, β NTI_{feat} has been adapted from β NTI to track the influence of individual features, which significantly contribute either to convergence or divergence.



Feature-level β -nearest taxon index can be calculated in three separate ways in order to account for variations in scale (e.g., a complete experiment vs. sub-groupings within the experiment). First, it can be calculated for an entire dataset yielding a single β NTI_{feat} value for each feature within a dataset. This approach is particularly useful in identifying which features in the whole dataset contribute to convergence or divergence. Second, it can be calculated within groups, for example, if you have a factor with two levels, you can calculate β NTI_{feat} for each of those levels. Results from this type of analysis will generate a β NTI_{feat} value for each feature within each level of the factor (i.e., if there are two factors and 100 features, 200 β NTI_{feat} values will be created) and is well-suited to compare differential feature contribution. Finally, β NTI_{feat} can be calculated truly pairwise (akin to the standard β NTI calculation) in order to provide complete spatial or temporal resolution at the feature level. By performing the β NTI_{feat} analysis on a fixed focal sample (a 0 day dry sample here),

you can get a true temporal perspective. Importantly, all these calculations can be done with absolute abundance, relative abundance, or presence/absence data.

The interpretation of the β NTI_{feat} is directly analogous to the interpretation of β NTI but focuses on individual “features” instead. Here, a “feature” is any member of a community or assemblage that is subject to ecological pressures (a microorganism or environmental metabolite, for example; **Table 1**). $|\beta$ NTI_{feat}| < 1 means that a feature has an insignificant contribution to ecological variation across the metacommunity or meta-assemblage, $1 < |\beta$ NTI_{feat}| < 2 indicates that a feature somewhat contributes ecological variation, and $|\beta$ NTI_{feat}| > 2 suggests that a given feature significantly contributes to ecological variation. These patterns can be further resolved based upon the sign of β NTI_{feat}. When β NTI_{feat} trends negative (e.g., < -1), we suggest that the feature contributes to convergence within the scale of analysis. Under this definition, these are features (or groups of related features) that significantly drive relational commonalities across a given

analytical scale. For example, these features could represent a phylogenetically conserved niche of microorganisms or consistent group of molecular formula that are disproportionately impacted by selective processes. When $\beta\text{NTI}_{\text{feat}}$ instead trends positive (e.g., >1), we suggest that the feature contributes to divergence. Under this definition, these are features (or groups of related features) that drive relational differences across an analytical scale. These could be those microorganisms which arose to prominence under varied environmental conditions (e.g., they were selected for/against) or those organic matter constituents that were produced/consumed under specific conditions.

Network Analysis

Weighted gene co-expression network analysis (WGCNA) was used to relate $\beta\text{NTI}_{\text{feat}}$ results and identify modules of related contributions to ecological assembly within community types (e.g., within the putative active community or molecular formula assemblage) or across community types (e.g., the putatively active community compared to the molecular formula assemblage; WGCNA package v1.70-3; Langfelder and Horvath, 2008). Networks were first generated in R and then visualized/analyzed using Cytoscape v3.8.2 (Shannon et al., 2003). Specifically, $\beta\text{NTI}_{\text{feat}}$ values for the putatively active ASVs (as identified *via* cDNA amplicon data) were related to the $\beta\text{NTI}_{\text{feat}}$ values for the molecular formulas. As part of this analysis, we identify modules of interconnected features (either ASVs or formula). These modules represent those features which have either positive or negative relationships in their contributions to ecological or functional convergence/divergence. Additional ecological metrics were calculated for modules containing both ASVs and molecular formulas including Shannon's diversity (H ; *diversity*, vegan package v2.5-7), $\exp(H)$ per recommendation from Jost (2006), Pielou's evenness (J), number of taxonomic Orders within a module, and the number of ASVs in a module (e.g., richness) The relative abundance of elemental composition groups was also calculated for these modules. The Cytoscape network file is stored as **Supplementary File 3**.

Data and Code Availability

Fourier transform ion cyclotron resonance mass spectrometer data are available on the ESS-DIVE archive at <https://data.ess-dive.lbl.gov/view/doi:10.15485/1807580>, sequence data are available on the NCBI Sequence Read Archive PRJNA641165, and scripts used throughout this manuscript (including various versions of the $\beta\text{NTI}_{\text{feat}}$ calculation) can be found on GitHub at <https://github.com/danczakre/betaNTI-feature> (Sengupta et al., 2021b).

RESULTS AND DISCUSSION

$\beta\text{NTI}_{\text{feat}}$ Revealed That Uncultured Microbial Lineages Differentially Contribute to Ecological Processes Across Total and Putatively Active Communities

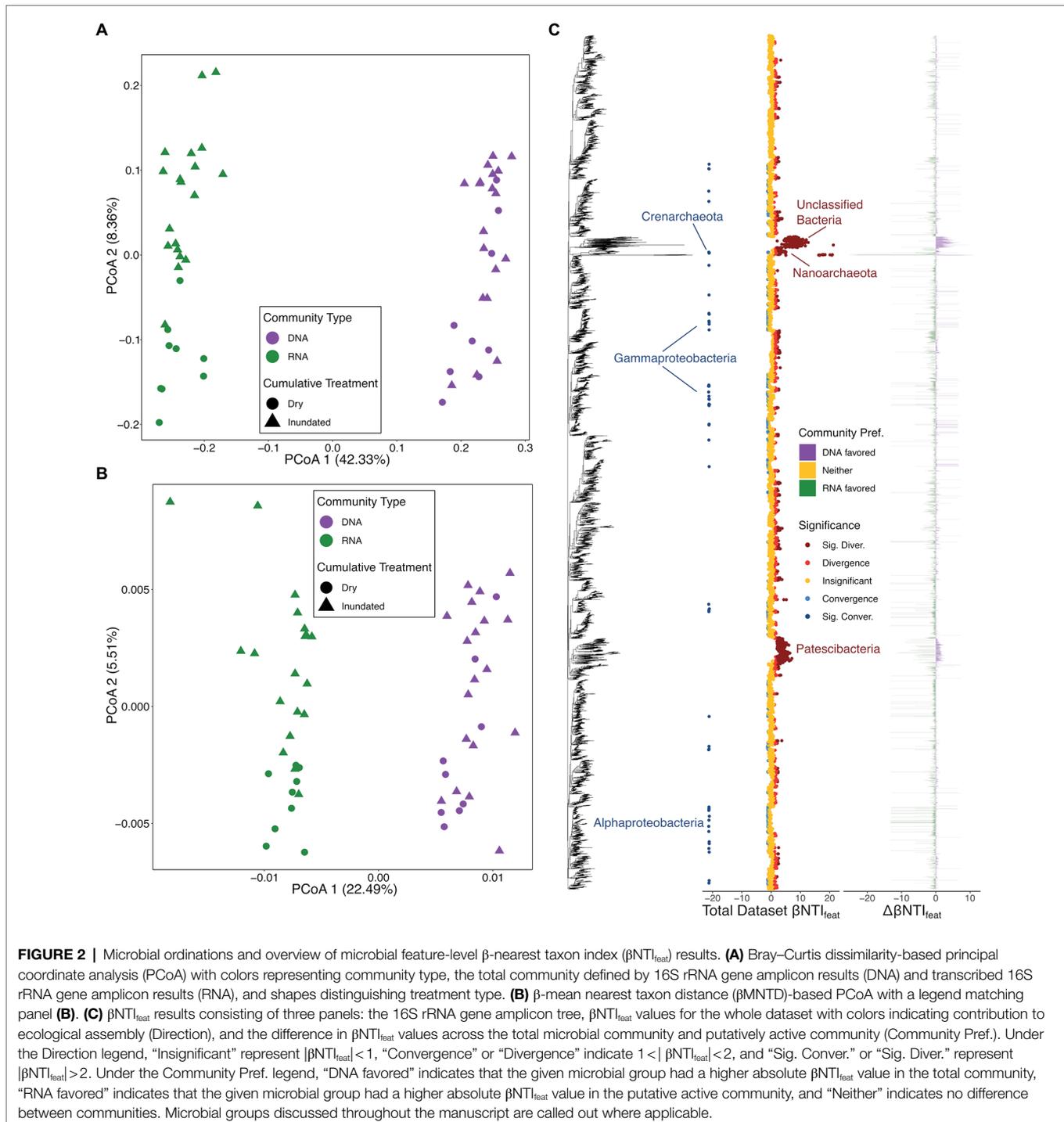
We first analyzed the overall microbial dynamics and observed that the total community (using 16S rRNA gene sequencing)

and putatively active community (*via* transcribed 16S rRNA sequencing) were significantly divergent from each other (**Figures 2A,B**). A Jaccard-based PCoA of Family-level taxonomic assignments revealed that not only were the two community types divergent, but also the communities under cumulatively dry treatments were divergent from those under inundated treatments (DNA – PERMANOVA Pseudo-F: 1.4826, value of $p < .05$; RNA – Pseudo-F: 3.4025, value of $p < .001$; and Type – PERMANOVA Pseudo-F: 38.653, value of $p < .001$; **Figure 2A**). Similar cross treatment dynamics were also apparent once sample similarity was analyzed using phylogenetic relatedness (e.g., βMNTD ; DNA – PERMANOVA Pseudo-F: 1.8191, value of $p < .05$; RNA – Pseudo-F: 3.7559, value of $p < .01$; and Type – PERMANOVA Pseudo-F: 69.54, value of $p < .001$; **Figure 2B**). Given that the largest differences existed between the total microbial community and the putatively active component (based upon Pseudo-F statistic for the Type comparison), this information was used as a baseline of differences to interpret downstream analyses.

When compared to existing taxonomic metrics (e.g., SIMPER), $\beta\text{NTI}_{\text{feat}}$ allows researchers to identify microbial groups impacting ecological structure rather than composition providing insight into community development (Clarke, 1993). In turn, we can evaluate the degree to which specific microbes contribute to the breadth of ecological strategies contained within metacommunities. This goes far beyond information on variation in taxonomic composition or standard diversity metrics. Existing approaches can identify cross-community differences in the abundance of a given taxon, but cannot identify the assembly processes leading to those differences in abundance. For each taxon (or molecule), $\beta\text{NTI}_{\text{feat}}$ quantifies the assembly processes it experiences, the degree to which it influences ecological structure, and elucidates the environmental conditions that lead to variation in these influences and contributions.

Feature-level β -nearest taxon index revealed that varied microbial lineages differentially impacted assembly of the total microbial community and the putatively active microbial community (**Figure 2C**). Looking at the $\beta\text{NTI}_{\text{feat}}$ dynamics across both the total and putative active communities, we see that members of an unclassified/uncultured group of Bacteria, members of the Patescibacteria (intermingled with some Class Bacilli), and members of the Nanoarchaeota consistently contribute to ecological divergence (**Figure 2C**). Those taxa which contribute to ecological convergence; however, are less conserved and more widespread across the phylogenetic tree with ASVs appearing within the Crenarchaeota, Alphaproteobacteria, and Gammaproteobacteria (**Figure 2C**).

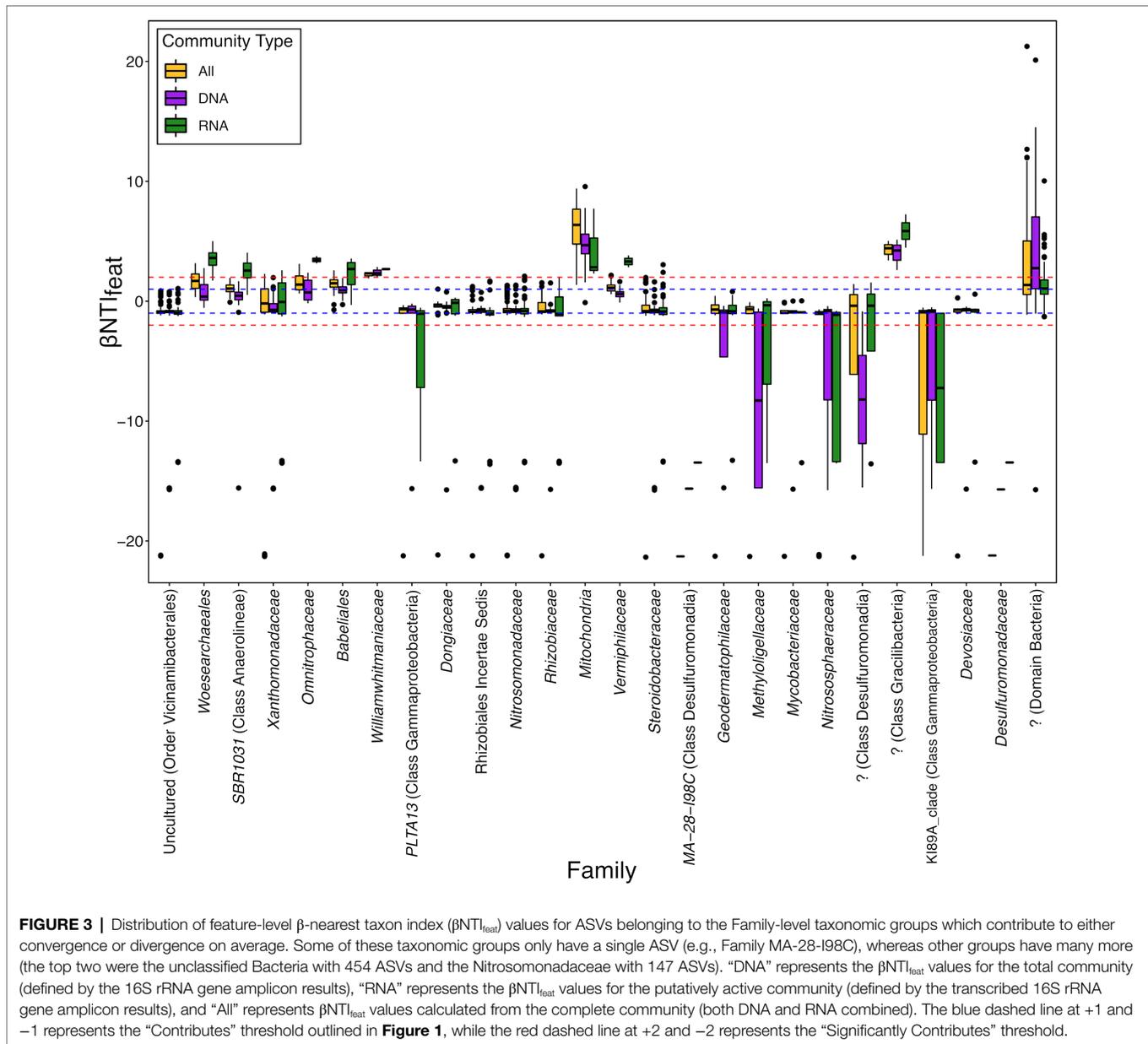
Given that the largest community differences existed between the total and putatively active communities (**Figures 2A,B**), we focused on evaluating which microbial groups differentially contributed across these groups. We see that those members of the uncultured bacterial group and the Patescibacteria are among the most notably differential. Specifically, while many members appear to contribute to either convergence or divergence across both communities, we see that fewer members of these lineages contribute to



assembly within the putatively active community. Out of 364 detected Patescibacteria across all microbial data, 305 contributed to assembly ($|\beta\text{NTI}_{\text{feat}}| > 1$) in the total community, while only 34 impacted the putatively active community (**Figure 2C**). This pattern demonstrates that the uncultured microbial lineages contribute more to the phylogenetic structure of the total community than the active community. The stronger role played by uncultured lineages in the total

community may indicate that these taxa have a background role and are not as relevant to the ecology of the active community.

Breaking the data down based upon Family-level taxonomic groups that, on average, contributed to either convergence or divergence supported these broad phylogenetic patterns and we observed that contributions could be related to inferred functional potentials and spanned a range of relative abundances

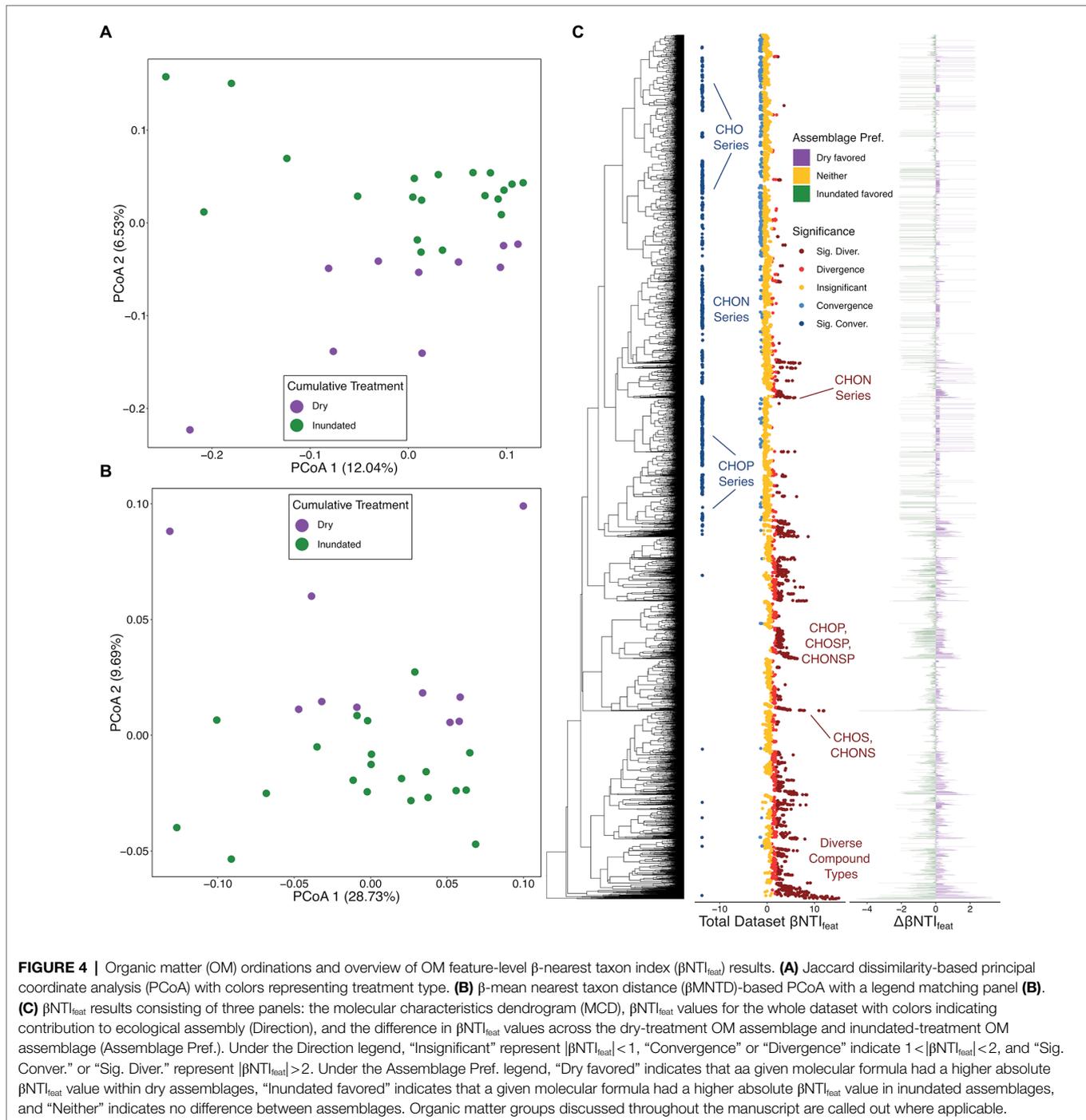


(**Figure 3; Supplementary Figure 1**). For example, an unknown family from Class Desulfuromonadia significantly drives convergence ($\beta\text{NTI}_{\text{feat}} < -2$) in the total community (DNA) but not in either the overall dataset (All) or putatively active (RNA) communities (**Figure 3**). Conversely, the Family *Woesearchaeales* significantly drives ecological divergence ($\beta\text{NTI}_{\text{feat}} > 2$) within the putatively active community as compared to the total community or overall dataset. In the case of these example taxonomic groups, we hypothesize that this is likely tied to their respective inferred functional potentials. Class Desulfuromonadia tend to feature sulfate reducing bacteria, while described members of Family *Woesearchaeales* are fermentative (Castelle et al., 2015; Waite et al., 2020). Given that the experimental design allowed these sediments to maintain oxic conditions, this would act as a selective force against sulfate reducers due to thermodynamic

constraints but not necessarily fermentative organisms (i.e., members could still form syntrophic relationships). This would translate to a diminished presence within the putative active community and thereby prevent them from significantly contributing to either convergence or divergence and suggests that these organisms might be detected in the DNA dataset as either dormant cells or relic DNA (Lennon et al., 2018).

Homologous Series Disproportionately Contribute to Convergence Regardless of Environment Type

The overall DOM patterns largely mirror those of the putatively active microbial community. Namely, both Jaccard- and βMNTD -based analyses revealed that significant differences existed



between the organic matter from dry and inundated samples (PERMANOVA Pseudo-F: 1.5085, value of $p < .05$; **Figures 4A,B**). Using this information, we calculated β NTI_{feat} for the combined dataset, the dry-only dataset, and the inundated dataset.

Looking first at the complete dataset, we see a higher proportion of features significantly contributing to functional divergence or convergence within the organic matter dataset (45.08%) than in the ASV dataset (20.99%). This potentially points simultaneously to the more transitive nature of many

environmental metabolites as compared to microbial community members (Schmidt et al., 2011; Graham et al., 2018; Coward et al., 2019), while also highlighting the consistent role that some conserved organic matter constituents play. Specifically, we see many larger molecular formula (e.g., C₄₇H₆₈O₁₀, C₃₄H₆₃O₄P, etc.) and some molecular formula with more complex compositions (e.g., C₄₆H₆₈NO₄S₂P, C₄₂H₅₂NO₉S₂P, etc.) significantly contribute to divergence across the entire dataset (**Figure 4C**). This might suggest that these more complex

compounds may play variable roles depending on the situation; they could be a hallmark of differential nutrient limitation (i.e., the functional needs of the meta-assemblage are highly variable; Graham et al., 2018; Garayburu-Caruso et al., 2020; Danczak et al., 2021) or highlight divergent degradation potential (e.g., common metabolites might be processed through different pathways).

Whereas divergence was driven by larger or more complex molecular formula compositions, we see that CHO-based homologous series (e.g., C₂₃H₃₀O₁₄, C₂₄H₃₀O₁₄, etc.), N-containing homologous series (e.g., C₁₈H₁₅NO₁₀, C₁₉H₁₇NO₁₀), and some P-containing homologous series (e.g., C₁₇H₂₅O₉P, C₁₈H₂₇O₉P) often contribute to convergence across the entire dataset. CHO-only molecular formulas are the most frequently observed types of organic matter from sedimentary and aquatic sources (Garayburu-Caruso et al., 2020) suggesting that these types of compounds could drive functional similarities across organic matter assemblages. As with those compounds which contribute to divergence, the contribution to convergence by the N-containing and P-containing formulas may also point to conserved functional needs or pathways (i.e., N/P-limitation drives selection for N/P-containing formulas, or ongoing metabolisms continually cycling these formulas).

Following the patterns established *via* the Jaccard- and β MNTD-based PCoAs, we evaluated the differences in environmental metabolites which contribute to convergence and divergence across the dry and inundated samples (Figure 4C). Unlike the ASV data, where groups contributed differently across the total and putatively active communities, we see a balanced set of contributions from formulas derived from dry and inundated organic matter with no clear discernable patterns.

Grouping molecular formula by either elemental composition categories or compound classes further revealed little consistent variation across wet-dry conditions suggesting that the OM meta-assemblage might be more impacted by ecosystem than treatment condition in this case (Figure 5). Similar patterns were observed in a headwater stream, where bulk environmental properties were conserved across ecosystem types despite divergent OM assemblages undergoing variable selection as a result of hypothesized thermodynamic redundancy (Danczak et al., 2021). Briefly, thermodynamic redundancy is like functional redundancy, where compositionally divergent OM assemblages have similar thermodynamic properties. Here, β NNTI_{feat} helps us examine whether this hypothesized redundancy exists across groups contributing to community structure. For example, in contrast to the thermodynamic behavior observed in Danczak et al. (2021), we see that the nominal oxidation state of carbon (NOSC) significantly varies across those CHON and CHOP containing molecular formula which contribute to convergence/divergence (Mann–Whitney U test value of $p < .05$; Supplementary Figure 2). Specifically, we observe that CHOP formulas with less inferred structural complexity (e.g., lower aromaticity index and double-bond equivalents) and more inferred lability (e.g., lower NOSC) drove significant convergence; differences in the properties of CHON formulas

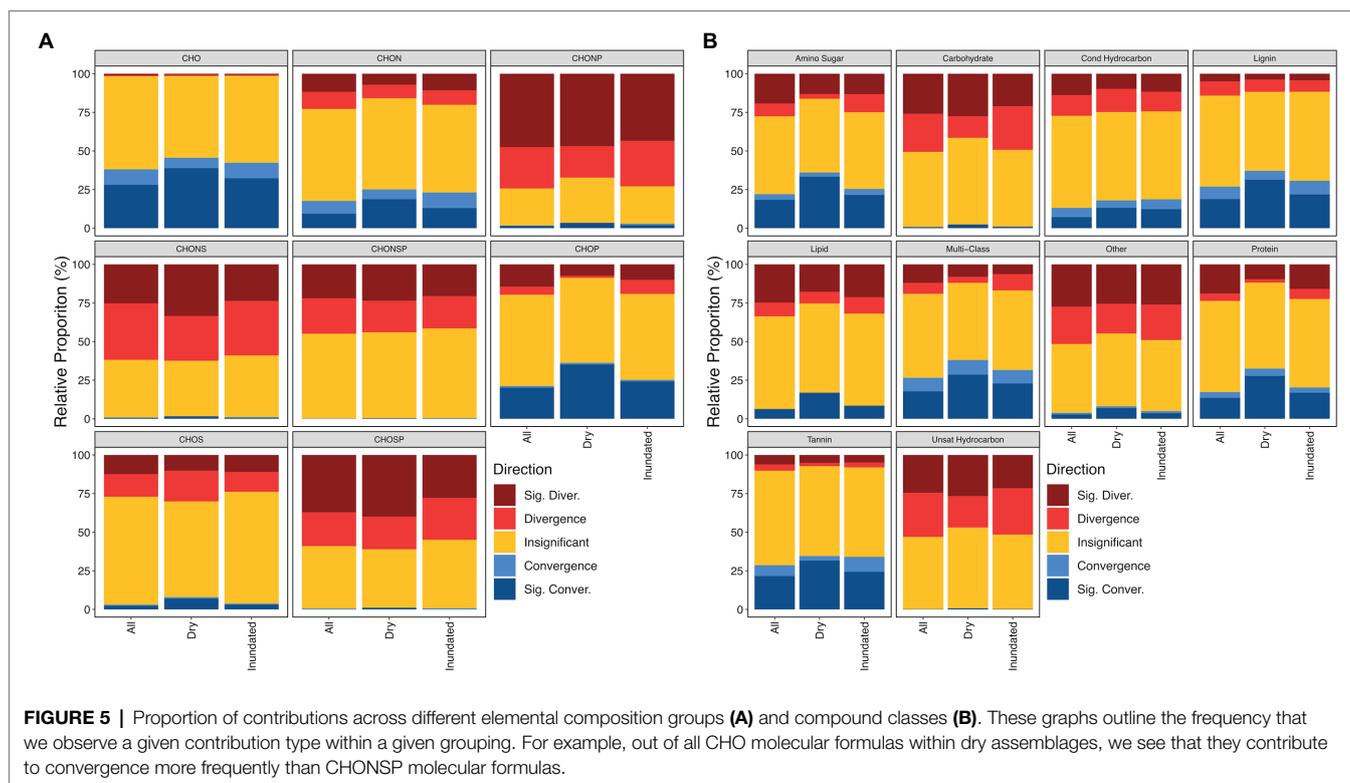
appear to manifest more as variation in distributions (e.g., those formulas which contribute to convergence are more constrained). Formulas featuring other elemental compositions exhibit less consistent behavior. This pattern suggests that thermodynamic restrictions potentially help dictate the contributions of the N/P-containing homologous series to the overall OM assemblage and may indicate that thermodynamic redundancy is not at play in this particular system. In other words, the preference for certain N/P-containing formulas might be dictated by organisms targeting the most preferential carbon source rather than some other limitation.

These results also demonstrate the capability for β NNTI_{feat} to uncover functional dynamics within OM assemblages and assign ecological importance in the absence of abundance information.

Network Analyses Reveal Groups of Molecular Formula and Microbes Which Contribute to Ecosystem Dynamics in a Coordinated Manner

While the previous sections focused on either analyzing complete datasets or subgroups within the overall dataset, β NNTI_{feat} also provides sample-level information for each feature within a given dataset (Supplementary Figure 3). Using this approach, we can evaluate how the contribution of each feature to community assembly changes through either time or space by utilizing a single focal sample (here ECA_0Cyc_R2). Looking at the five most variable taxonomic groups from the putatively active community, we can see that different groups respond differently through treatment regimes (Supplementary Figure 3). For example, families DTB120, MVP-15, and order Myxococcales rarely if ever cross the +2 β NNTI_{feat} threshold for contributing to divergence. In contrast, the families *Williamwhitmaniaceae* and *Woesearchaeales* consistently cross that threshold. In contrast to these highly dynamic ASV taxa, the groups defined by a molecular formula's elemental composition are less variable across samples (Supplementary Figure 3).

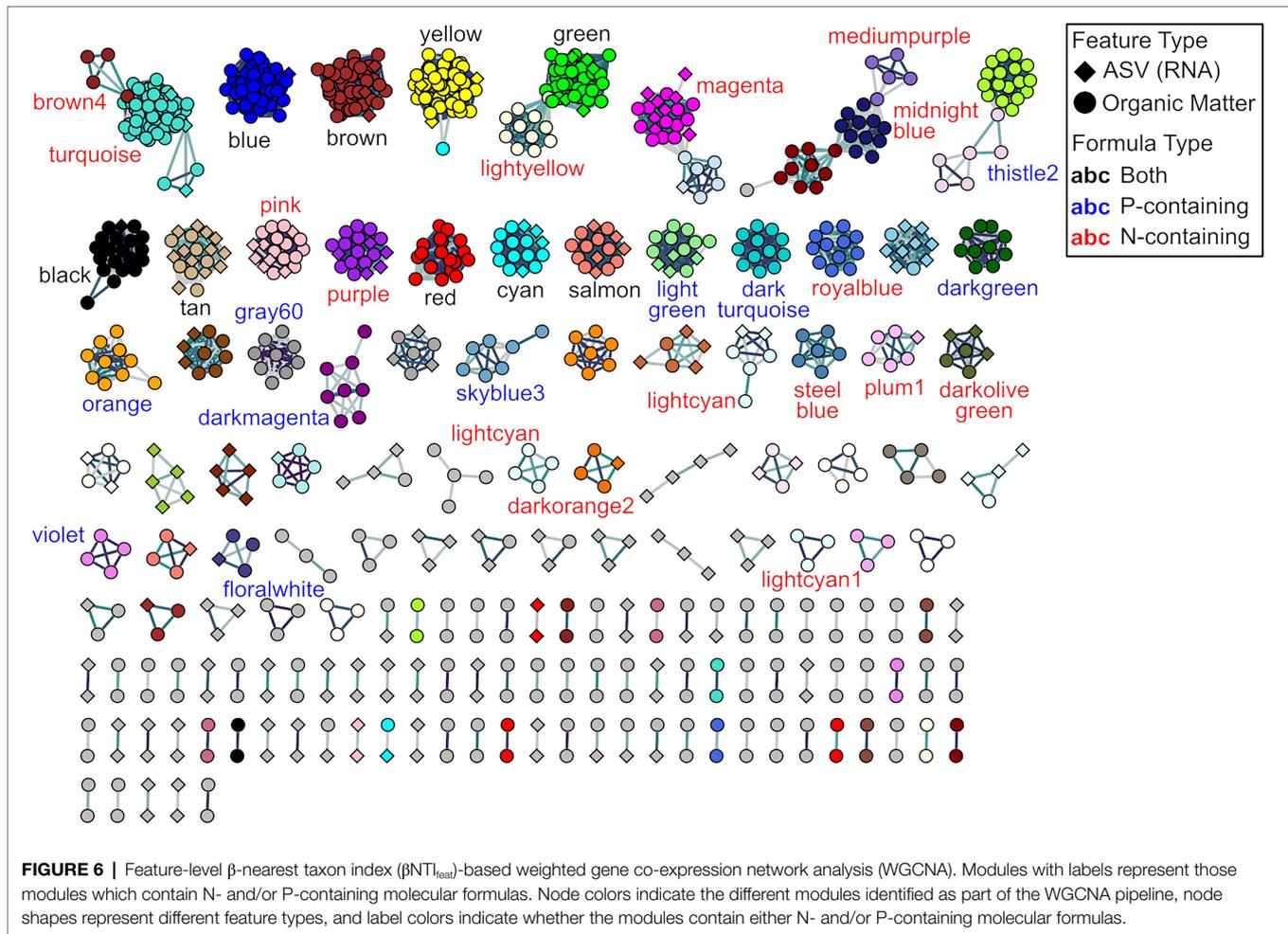
Having established that β NNTI_{feat} can track dynamics through time, we believe that this metric is well-suited in network analyses. By performing a WGCNA using β NNTI_{feat} values for each feature—as opposed to abundances—we can obtain modules of putatively active ASVs and OM formulas. These modules should reflect features which have either positive or negative relationships among their contributions to ecological/functional convergence and divergence. Within-cluster relationships between ASVs and organic molecules do not indicate co-variance in abundances (as in a traditional network analysis). Instead, these relationships indicate linkages between the ecology of microbes and the functional properties of organic molecules. We propose that these putative ecology-function linkages are more likely to indicate causal connections between microbes and molecules than networks built on relative abundances. There are nonetheless multiple potential interpretations of within-cluster relationships, leading to additional questions. For example, could the ASVs be responsible for degrading the linked molecular formulas?



If a module contains exclusively molecular formulas, does this point a single pathway or do they represent an ecological interaction among pathways? By revealing features that have coordinated impacts on the ecological or functional structure of their respective community, network analyses can highlight important organisms or molecules. A better understanding of the relationships among ecological pressures impacting members across communities within a metacommunity will enable researchers to develop transitive principles regarding the coordination of ecological assembly (e.g., the mechanisms impacting cross trophic level assembly, etc.).

The network analysis resulted in the creation of 54 modules, with 28 modules having some mixture of ASVs and organic matter, two featuring exclusively ASVs, and 24 containing only molecular formulas (Figure 6; Supplementary Figure 4). Of these 54 modules, two modules were removed because they consisted only of doublets (e.g., modules where one feature related to only one other single feature). Focusing on the 27 modules which have a mixture of ASVs and molecular formulas, we observe a range of microbial diversity: for example, we observe that the gray and tan modules have the greatest number of different taxonomic groups (Supplementary Figure 4A). These modules also exhibit a range of characteristics across their molecular formulas: we see that the gray module predominantly contains CHO-only formulas, while the tan module primarily contains CHON formulas (Supplementary Figure 4B). This approach now allows us to evaluate which factors and taxonomic groups are coordinated in driving ecological dynamics.

Given the importance of nitrogen and phosphorus homologous uncovered in the metabolite β NTI_{feat} analyses, we examined modules containing at least two N- or P-containing formulas and contained more than two nodes. There were 16 modules with two or more N-containing molecular formulas, 10 modules with two or more P-containing formulas, and nine modules with both two or more N- and P-containing formulas (Figure 6). Analyzing the membership of these modules first revealed that the P-containing modules only featured four ASVs across all 10 modules, with three modules comprised exclusively of P-containing modules. This pattern suggests that these P-containing molecular formulas potentially experience a set of selective pressures distinct from other metabolites and ASVs. This characteristic could explain the disproportionate effect that phosphorus homologous series had on the functional structure of the OM meta-assemblage. In contrast, N-containing modules featured broader microbial and molecular diversity indicating that N-containing molecular formulas experience a more common set of pressures than P-containing formulas. As hypothesized above, this might be driven by nutritional requirements or some sort of pathway regulation. Finally, five different ASVs from Family *Geobacteraceae* were detected in three out of the nine modules featuring both N- and P-containing molecular formulas. While we do not have direct functional information due to the limits of amplicon sequencing, members of this family have reported roles in both the nitrogen cycle (*via* nitrate reduction, nitrification, and nitrogen fixation) and the phosphorus cycle (*via* aggressive phosphate acquisition; Naik et al., 1993; N'Guessan et al., 2010; Ueki and Lovley, 2010;



Wang et al., 2018; Samaddar et al., 2019; Campeciño et al., 2020). Taken together, this potentially points to a combinatorial effect, whereby the functional development driven by N/P-containing formulas is tied to organismal selectivity (i.e., roles of *Geobacteraceae* in the N/P cycles).

CONCLUSION

Feature-level β -nearest taxon index is a novel metric that enables researchers to investigate the contributions to ecological convergence and divergence with a given metacommunity or meta-assemblage. Using $\beta\text{NTI}_{\text{feat}}$, we revealed that uncultured lineages contribute significantly to the ecological structure of microbial communities when assayed using 16S rDNA, but contributed little to the ecology of putatively active communities assayed with 16S rRNA. These differences suggest that while these uncultured lineages represent a significant proportion of the microorganisms in some ecosystems (Brown et al., 2015; Anantharaman et al., 2016; Danczak et al., 2017; León-Zayas et al., 2017), they may be relatively minor contributors to the ecological composition of active microbes in the study system. This inference is complementary to traditional inferences based

on relative abundances or taxonomic richness. That is, $\beta\text{NTI}_{\text{feat}}$ quantifies the contribution of individual ASVs to the assembly of multi-dimensional ecological niche space occupied by communities, and how those contributions vary through space, time, and environmental conditions.

Applying $\beta\text{NTI}_{\text{feat}}$ to OM provides analogous inferences, but instead quantifies the contribution of individual molecules to the assembly of the multi-dimensional functional space occupied by assemblages of organic molecules. In our study, we observed that OM homologous series primarily containing nitrogen or phosphorus differentially contributed to OM functional composition across the whole dataset. However, we do not see any noteworthy differences across wet-dry treatments indicating a stronger ecosystem-level control. Network analyses relating putatively active ASV $\beta\text{NTI}_{\text{feat}}$ values to metabolite $\beta\text{NTI}_{\text{feat}}$ values revealed formation of 54 distinct modules exhibiting coordinated contributions to their respective communities. Looking specifically at modules featuring N/P-containing molecular formulas, we observed that P-containing formulas appear to be largely disconnected from other metabolites and microorganisms, N-containing formulas are more integrated across feature types, and modules with both formula types exhibit a high incidence of *Geobacteraceae* members.

These dynamics potentially point to two larger scale hypotheses. First, the ecological contributions of P-containing formulas are often distinct from other metabolites indicating that they might occupy a unique functional space (at least compared to N-containing formulas). Second, certain microbial groups may offer contributions to convergence or divergence in concert with specific metabolite dynamics (here *Geobacteraceae* related to N/P-containing formulas). While these hypotheses still need to be independently verified, they point to specific conclusions that can be drawn from the use of $\beta\text{NTI}_{\text{feat}}$. As further research uses this metric, more broad scale, transferrable hypotheses can be generated and help us better understand community assembly.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: FTICR-MS data are available on the ESS-DIVE archive at: <https://data.ess-dive.lbl.gov/view/doi:10.15485/1807580> and sequence data are available on the NCBI Sequence Read Archive PRJNA641165.

AUTHOR CONTRIBUTIONS

RD and JS conceptualized and designed the study. AS assisted in experimental design. JT, JW, LR, RC, SF, and VG-C performed

the experiment. RD analyzed the data. RD and JS drafted the manuscript and all authors contributed to further writing. All authors contributed to the article and approved the submitted version.

FUNDING

The initial experimental stages of this work were supported by the PREMIS Initiative at the Pacific Northwest National Laboratory (PNNL) with funding from the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the United States Department of Energy under Contract DE-AC05-76RL01830. The later stages of this work (e.g., data analysis, conceptual interpretation manuscript development) were supported by the United States Department of Energy-BER program, as part of an Early Career Award to JS at PNNL. A portion of the research was performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.803420/full#supplementary-material>

REFERENCES

- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., et al. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* 7:13219. doi: 10.1038/ncomms13219
- Bailey, V. L., Smith, A. P., Tfaily, M., Fansler, S. J., and Bond-Lamberty, B. (2017). Differences in soluble organic carbon chemistry in pore waters sampled from different pore size domains. *Soil Biol. Biochem.* 107, 133–143. doi: 10.1016/j.soilbio.2016.11.025
- Blomberg, S. P., Garland, T., and Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57, 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x
- Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90. doi: 10.1186/s40168-018-0470-z
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9
- Bottos, E. M., Kennedy, D. W., Romero, E. B., Fansler, S. J., Brown, J. M., Bramer, L. M., et al. (2018). Dispersal limitation and thermodynamic constraints govern spatial structure of permafrost microbial communities. *FEMS Microbiol. Ecol.* 94:fiy110. doi: 10.1093/femsec/fiy110
- Bramer, L. M., White, A. M., Stratton, K. G., Thompson, A. M., Claborne, D., Hofmockel, K., et al. (2020). FtmsRanalysis: an R package for exploratory data analysis and interactive visualization of FT-MS data. *PLoS Comput. Biol.* 16:e1007654. doi: 10.1371/journal.pcbi.1007654
- Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., et al. (2015). Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* 523, 208–211. doi: 10.1038/nature14486
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Campeciño, J., Lagishetty, S., Wawrzak, Z., Alfaro, V. S., Lehnert, N., Reguera, G., et al. (2020). Cytochrome c nitrite reductase from the bacterium *Geobacter lovleyi* represents a new NrfA subclass. *J. Biol. Chem.* 295, 11455–11465. doi: 10.1074/jbc.RA120.013981
- Castelle, C. J., Wrighton, K. C., Thomas, B. C., Hug, L. A., Brown, C. T., Wilkins, M. J., et al. (2015). Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25, 690–701. doi: 10.1016/j.cub.2015.01.014
- Chase, J. M., and Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2351–2363. doi: 10.1098/rstb.2011.0063
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Austral. Ecol.* 18, 117–143. doi: 10.1111/j.1442-9993.1993.tb00438.x
- Coward, E. K., Ohno, T., and Sparks, D. L. (2019). Direct evidence for temporal molecular fractionation of dissolved organic matter at the iron oxyhydroxide interface. *Environ. Sci. Technol.* 53, 642–650. doi: 10.1021/acs.est.8b04687
- Danczak, R. E., Chu, R. K., Fansler, S. J., Goldman, A. E., Graham, E. B., Tfaily, M. M., et al. (2020a). Using metacommunity ecology to understand environmental metabolomes. *Nat. Commun.* 11:6369. doi: 10.1038/s41467-020-19989-y
- Danczak, R. E., Daly, R. A., Borton, M. A., Stegen, J. C., Roux, S., Wrighton, K. C., et al. (2020b). Ecological assembly processes are coordinated between bacterial and viral communities in fractured shale ecosystems. *mSystems* 5, e00098–e00120. doi: 10.1128/mSystems.00098-20
- Danczak, R. E., Goldman, A. E., Chu, R. K., Toyoda, J. G., Garayburu-Caruso, V. A., Tolić, N., et al. (2021). Ecological theory applied to environmental metabolomes reveals compositional divergence despite conserved molecular properties. *Sci. Total Environ.* 788:147409. doi: 10.1016/j.scitotenv.2021.147409

- Danczak, R. E., Johnston, M. D., Kenah, C., Slattery, M., Wrighton, K. C., and Wilkins, M. J. (2017). Members of the candidate phyla radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome* 5:112. doi: 10.1186/s40168-017-0331-1
- Dittmar, T., Koch, B., Hertkorn, N., and Kattner, G. (2008). A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* 6, 230–235. doi: 10.4319/lom.2008.6.230
- Fodelianakis, S., Washburne, A. D., Bourquin, M., Pramatefaki, P., Kohler, T. J., Styllas, M., et al. (2021). Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. *ISME J.* doi:10.1038/s41396-021-01106-6 [Epub ahead of print].
- Garayburu-Caruso, V. A., Danczak, R. E., Stegen, J. C., Renteria, L., Mccall, M., Goldman, A. E., et al. (2020). Using community science to reveal the global chemogeography of river metabolomes. *Metabolites* 10:518. doi: 10.3390/metabo10120518
- George, D. B., Webb, C. T., Farnsworth, M. L., O'Shea, T. J., Bowen, R. A., Smith, D. L., et al. (2011). Host and viral ecology determine bat rabies seasonality and maintenance. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10208–10213. doi: 10.1073/pnas.1010875108
- Gilbert, B., and Bennett, J. R. (2010). Partitioning variation in ecological communities: do the numbers add up? *J. Appl. Ecol.* 47, 1071–1082. doi: 10.1111/j.1365-2664.2010.01861.x
- Graham, E. B., Crump, A. R., Kennedy, D. W., Arntzen, E., Fansler, S., Purvine, S. O., et al. (2018). Multi 'omics comparison reveals metabolome biochemistry, not microbiome composition or gene expression, corresponds to elevated biogeochemical function in the hyporheic zone. *Sci. Total Environ.* 642, 742–753. doi: 10.1016/j.scitotenv.2018.05.256
- Herren, C. M., and McMahon, K. D. (2017). Cohesion: a method for quantifying the connectivity of microbial communities. *ISME J.* 11, 2426–2438. doi: 10.1038/ismej.2017.91
- Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G., and Qian, K. (2001). Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* 73, 4676–4681. doi: 10.1021/ac1010560w
- Jiao, S., Yang, Y., Xu, Y., Zhang, J., and Lu, Y. (2020). Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J.* 14, 202–216. doi: 10.1038/s41396-019-0522-9
- Jost, L. (2006). Entropy and diversity. *Oikos* 113, 363–375. doi: 10.1111/j.2006.0030-1299.14714.x
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., et al. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464. doi: 10.1093/bioinformatics/btq166
- Kim, S., Kramer, R. W., and Hatcher, P. G. (2003). Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van krevelen diagram. *Anal. Chem.* 75, 5336–5344. doi: 10.1021/ac034415p
- Koch, B. P., and Dittmar, T. (2006). From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. *Rapid Commun. Mass Spectrom.* 20, 926–932. doi: 10.1002/rcm.2386
- Kraft, N. J. B., Cornwell, W. K., Webb, C. O., and Ackerly, D. D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *Am. Nat.* 170, 271–283. doi: 10.1086/519400
- Kujawinski, E. B., and Behn, M. D. (2006). Automated analysis of electrospray ionization fourier transform ion cyclotron resonance mass spectra of natural organic matter. *Anal. Chem.* 78, 4363–4373. doi: 10.1021/ac0600306
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- LaRowe, D. E., and Van Cappellen, P. (2011). Degradation of natural organic matter: a thermodynamic analysis. *Geochim. Cosmochim. Acta* 75, 2030–2042. doi: 10.1016/j.gca.2011.01.020
- Lennon, J. T., Muscarella, M. E., Placella, S. A., and Lehmkuhl, B. K. (2018). How, when, and where relic DNA affects microbial diversity. *mBio* 9, e00637–e00718. doi: 10.1128/mBio.00637-18
- León-Zayas, R., Peoples, L., Biddle, J. F., Podell, S., Novotny, M., Cameron, J., et al. (2017). The metabolic potential of the single cell genomes obtained from the challenger deep, mariana trench within the candidate superphylum Parcubacteria (OD1). *Environ. Microbiol.* 19, 2769–2784. doi: 10.1111/1462-2920.13789
- Moroenyane, I., Mendes, L., Tremblay, J., Tripathi, B., and Yergeau, E. (2021). Plant compartments and developmental stages modulate the balance between niche-based and neutral processes in soybean microbiome. *Microb. Ecol.* 82, 416–428. doi: 10.1007/s00248-021-01688-w
- N'Guessan, A. L., Elifantz, H., Nevin, K. P., Mouser, P. J., Methé, B., Woodard, T. L., et al. (2010). Molecular analysis of phosphate limitation in Geobacteraceae during the bioremediation of a uranium-contaminated aquifer. *ISME J.* 4, 253–266. doi: 10.1038/ismej.2009.115
- Naik, R. R., Francisco, M. M., and Stolz, J. F. (1993). Evidence for a novel nitrate reductase in the dissimilatory iron-reducing bacterium *Geobacter metallireducens*. *FEMS Microbiol. Lett.* 106, 53–58. doi: 10.1111/j.1574-6968.1993.tb05934.x
- Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X., et al. (2020). A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nat. Commun.* 11:4717. doi: 10.1038/s41467-020-18560-z
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2019). Vegan: Community Ecology Package. Available at: <https://cran.r-project.org/package=vegan> (Accessed August 15, 2021).
- Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219
- Rivas-Ubach, A., Liu, Y., Bianchi, T. S., Tolić, N., Jansson, C., and Paša-Tolić, L. (2018). Moving beyond the van krevelen diagram: a new stoichiometric approach for compound classification in organisms. *Anal. Chem.* 90, 6152–6160. doi: 10.1021/acs.analchem.8b00529
- Samaddar, S., Chatterjee, P., Truu, J., Anandham, R., Kim, S., and Sa, T. (2019). Long-term phosphorus limitation changes the bacterial community structure and functioning in paddy soils. *Appl. Soil Ecol.* 134, 111–115. doi: 10.1016/j.apsoil.2018.10.016
- Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., et al. (2011). Persistence of soil organic matter as an ecosystem property. *Nature* 478, 49–56. doi: 10.1038/nature10386
- Sengupta, A., Fansler, S. J., Chu, R. K., Danczak, R. E., Garayburu-Caruso, V. A., Renteria, L., et al. (2021a). Disturbance triggers non-linear microbe-environment feedbacks. *Biogeosciences* 18, 4773–4789. doi: 10.5194/bg-18-4773-2021
- Sengupta, A., Fansler, S. J., Toyoda, J. G., Chu, R. K., Garayburu-Caruso, V. A., Renteria, L., et al. (2021b). Respiration data, microbial community assembly data, and FTICR-MS data associated with: "Disturbance Triggers Non-Linear Microbe-Environment Feedbacks. Sengupta et al., 2021." *Biogeosciences* doi: 10.15485/1807580
- Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Smith, T. W., and Lundholm, J. T. (2010). Variation partitioning as a tool to distinguish between niche and neutral processes. *Ecography* 33, 648–655. doi: 10.1111/j.1600-0587.2009.06105.x
- Stegen, J. C., Fredrickson, J. K., Wilkins, M. J., Konopka, A. E., Nelson, W. C., Arntzen, E. V., et al. (2016). Groundwater-surface water mixing shifts ecological assembly processes and stimulates organic carbon turnover. *Nat. Commun.* 7:11237. doi: 10.1038/ncomms11237
- Stegen, J. C., Johnson, T., Fredrickson, J. K., Wilkins, M. J., Konopka, A. E., Nelson, W. C., et al. (2018). Influences of organic carbon speciation on hyporheic corridor biogeochemistry and microbial ecology. *Nat. Commun.* 9:585. doi: 10.1038/s41467-018-02922-9
- Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., et al. (2013). Quantifying community assembly processes and identifying features that impose them. *ISME J.* 7, 2069–2079. doi: 10.1038/ismej.2013.93

- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6, 1653–1664. doi: 10.1038/ismej.2012.22
- Swenson, N. G., Enquist, B. J., Pither, J., Thompson, J., and Zimmerman, J. K. (2006). The problem and promise of scale dependency in community phylogenetics. *Ecology* 87, 2418–2424. doi: 10.1890/0012-9658(2006)87[2418:TPAPOS]2.0.CO;2
- Tolić, N., Liu, Y., Liyu, A., Shen, Y., Tfaily, M. M., Kujawinski, E. B., et al. (2017). Formularity: software for automated formula assignment of natural and other organic matter from ultrahigh-resolution mass spectra. *Anal. Chem.* 89, 12659–12665. doi: 10.1021/acs.analchem.7b03318
- Ueki, T., and Lovley, D. R. (2010). Novel regulatory cascades controlling expression of nitrogen-fixation genes in *Geobacter sulfurreducens*. *Nucleic Acids Res.* 38, 7485–7499. doi: 10.1093/nar/gkq652
- Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., et al. (2020). Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* 70, 5972–6016. doi: 10.1099/ijsem.0.004213
- Wang, S., An, J., Wan, Y., Du, Q., Wang, X., Cheng, X., et al. (2018). Phosphorus competition in bioinduced vivianite recovery from wastewater. *Environ. Sci. Technol.* 52, 13863–13870. doi: 10.1021/acs.est.8b03022
- Warton, D. I., Wright, S. T., and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* 3, 89–101. doi: 10.1111/j.2041-210X.2011.00127.x
- Zhang, K., Shi, Y., Cui, X., Yue, P., Li, K., Liu, X., et al. (2019). Salinity is a key determinant for soil microbial communities in a desert ecosystem. *mSystems* 4, e00225–e00318. doi: 10.1128/msystems.00225-18
- Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81, e00002–e00017. doi: 10.1128/MMBR.00002-17

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Danczak, Sengupta, Fansler, Chu, Garayburu-Caruso, Renteria, Toyoda, Wells and Stegen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.