**Interpreting UniFrac with Absolute Abundance: A Conceptual and Practical Guide**

Augustus Pendleton1\* & Marian L. Schmidt1\*

1Department of Microbiology, Cornell University, 123 Wing Dr, Ithaca, NY 14850, USA

**Corresponding Authors:** Augustus Pendleton: arp277@cornell.edu; Marian L. Schmidt: marschmi@cornell.edu

**Author Contribution Statement:** Both authors contributed equally to the manuscript.

**Preprint servers:** This article was submitted to *bioRxiv* (doi: 10.1101/2025.07.18.665540) under a CC-BY-NC-ND 4.0 International license.

**Keywords:** Microbial Ecology - Beta Diversity - Absolute Abundance - Bioinformatics - UniFrac

**Data Availability:** All data and code used to produce the manuscript are available at https://github.com/MarschmiLab/Pendleton\_2025\_Absolute\_Unifrac\_Paper, in addition to a reproducible renv environment. All packages used for analysis are listed in Table S1.

**Abstract**

-diversity is central to microbial ecology, yet commonly used metrics overlook changes in microbial load, limiting their ability to detect ecologically meaningful shifts. Popular for incorporating phylogenetic relationships, UniFrac distances currently default to relative abundance and therefore omit important variation in microbial abundances. As methods for quantifying absolute abundance become more accessible, integrating this information into -diversity analyses becomes increasingly important. Here, we introduce *Absolute UniFrac* (), a variant of Weighted UniFrac that operates on absolute abundances. Using simulations and reanalysis of four 16S rRNA datasets (nuclear reactor cooling tank, the mouse gut, a freshwater lake, and the peanut rhizospere), we demonstrate that Absolute UniFrac jointly captures microbial load differences and phylogenetic turnover in comparison to other abundance-aware -diversity metrics. While this can improve statistical power to detect ecological shifts, Absolute Unifrac can also be strongly correlated to differences in cell abundances alone, especially when changes occur along long branches. To balance these effects, we present a generalized extension () with a tunable parameter to balance sensitivity and interpretability. We recommend evaluating conclusions across multiple parameters and complementary -diversity metrics. Finally, we benchmark *GUA* and show that although computationally slower than conventional alternatives, *GUA* is comparably insensitive to realistic noise in load estimates compared to conventional alternatives (i.e. Bray-Curtis dissimilarities), particularly at lower By coupling phylogeny with microbial load, Absolute UniFrac reframes -diversity as an eco-evolutionary measure of how communities reorganize in both membership and magnitude.

**Introduction**

Microbial ecologists routinely compare communities using -diversity metrics derived from relative abundances. Yet this approach overlooks a critical ecological dimension: microbial load. High-throughput sequencing produces compositional data, in which each taxon’s abundance is constrained by all others [1]. However, quantitative profiling studies show that cell abundance, not only composition, can drive major community differences [2]. In low-biomass samples, relying on relative abundance can allow contaminants to appear biologically meaningful despite absolute counts too low for concern [3].

Sequencing-based microbiome studies therefore rely on relative abundance even when the hypotheses of interest implicitly concern absolute changes in biomass. This creates a mismatch between ecological framing (growth, bloom magnitude, pathogen proliferation, disturbance recovery, or colonization pressure) and the information the -diversity metric encodes. As a result, -diversity is often treated as if it includes biomass, even when absolute abundance is either not measured at all or is measured but excluded from the calculation (as in conventional UniFrac). Many studies therefore test biomass-linked hypotheses using a metric that normalizes biomass away. A conceptual correction is needed in which -diversity is understood as variation along three axes: composition, phylogeny and biomass.

Absolute microbial load measurements are now increasingly obtainable through flow cytometry, qPCR, and genomic spike-ins, allowing biomass to be quantified alongside taxonomic composition [4, 5]. These approaches improve detection of functionally relevant taxa and mitigate the compositional constraints imposed by sequencing [1, 2]. Most studies that incorporate absolute counts currently rely on Bray-Curtis dissimilarity, which captures load but not phylogenetic turnover[4, 6]. UniFrac provides the opposite strength, incorporating phylogeny but discards absolute abundance as it is restricted to relative abundance data by construction [7]. This leaves no phylogenetically informed -diversity metric that operates on absolute counts, despite the fact that biomass is central to many ecological hypotheses.

To evaluate the implications of incorporating absolute abundance into phylogenetic -diversity, we use both simulations and reanalysis of four 16S rRNA amplicon datasets. The simulations use a simple four-taxon community with controlled abundance shifts to directly compare both Bray-Curtis and UniFrac in their relative and absolute forms, illustrating how each metric responds when abundance, composition, or evolutionary relatedness differ. We then reanalyze four real-world datasets spanning nuclear reactor cooling water [4], the mouse gut [3], a freshwater lake [10], and rhizosphere soil [11]. These differ widely in richness, biomass range, and ecological context, allowing us to test when absolute abundance changes align with or diverge from phylogenetic turnover. Together, these simulations and reanalyses provide the empirical foundation for interpreting Absolute UniFrac relative to existing -diversity measures across the three axes of ecological difference: abundance, composition, and phylogeny. Absolute UniFrac is a direct extension of Generalized Weighted UniFrac [7] that incorporates total abundance as a tunable ecological dimension, and evaluate its impact across simulated and real-world datasets. When = 0, it behaves like conventional unweighted UniFrac, reflecting purely compositional and phylogenetic turnover. As increases, it additionally captures abundance-driven divergence. This extension enables tree-based -diversity to operate in the absolute abundance domain, aligning the metric with ecological hypotheses in that depend on biomass change.

**Defining Absolute UniFrac**

The UniFrac distance was first introduced by Lozupone & Knight (2005) and has since become enormously popular as a measure of -diversity within the field of microbial ecology [7]. A benefit of the UniFrac distance is that is considers phylogenetic information when estimating the distance between two communities. After first generating a phylogenetic tree representing species (or amplicon sequence variants, “ASVs”) from all samples, the UniFrac distance computes the fraction of branch-lengths which is *shared* between communities, relative to the total branch length represented in the tree. UniFrac can be both “unweighted”, in which only the incidence of species is considered, or “weighted”, wherein a branch’s contribution is weighted by the proportional abundance of taxa on that branch [8]. The Weighted UniFrac is derived:

where the length of each branch is weighted, , by the difference in the relative abundance of all species () descended from that branch in sample or sample . Here, we denote this distance as *UR*, for “Relative UniFrac”. Popular packages which calculate weighted Unifrac, including the diversity-lib QIIME plug-in and the R packages phyloseq and GUniFrac, run this normalization by default.

Importantly, is most sensitive to changes in abundant lineages, which can sometimes obscure compositional differences driven by rare to moderately-abundant taxa [9]. To address this weakness, Chen et al. (2012) introduced the generalized UniFrac distance (), in which the impact of abundant lineages can be mitigated by decreasing the parameter :

where ranges from 0 (close to Unweighted UniFrac) up to 1 (identical to , above). However, if one wishes to use absolute abundances, both and can be derived without normalizing to proportions:

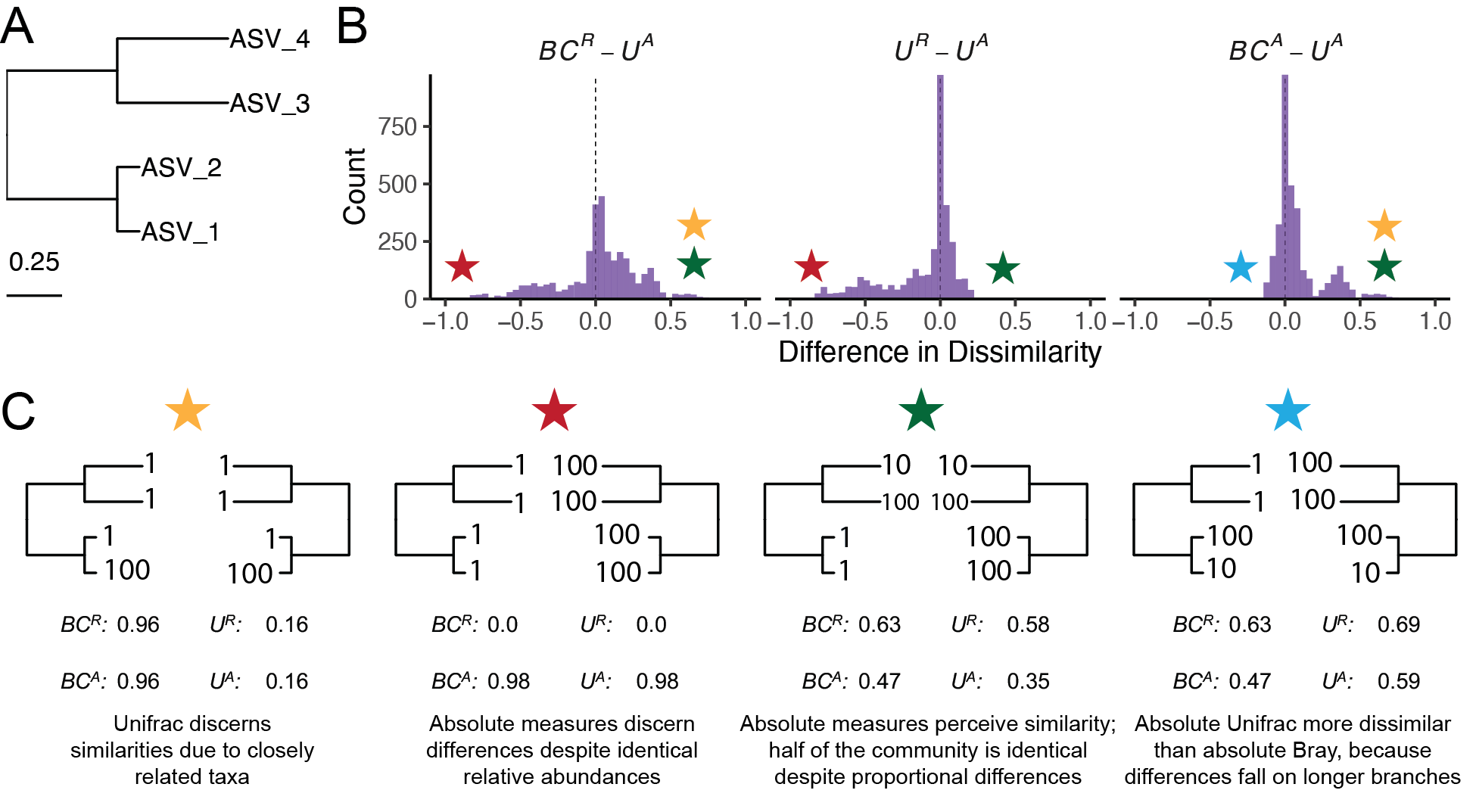
Where and denote the absolute counts of species descending from branch in communities and , respectively. We refer to these distances as *Absolute UniFrac* (). Although substituting absolute for relative abundances is mathematically straightforward, we found no prior work that examines UniFrac in the context of absolute abundance, either conceptually or in application. Incorporating absolute abundances introduces a third axis of ecological variation: beyond differences in composition and phylogenetic turnover, also captures divergence in microbial load. This has made interpretation of nontrivial, particularly in complex microbiomes, and has likely contributed to systematic under-interpretation of abundance-driven ecological processes in tree-based -diversity analyses.

**Demonstrating -diversity metrics behavior with a simple simulation**

To clarify how behaves relative to existing -diversity metrics, we first constructed a simple four-taxon simulated community arranged in a small phylogeny (Fig. 1A). By varying the absolute abundance of each taxon or ASV (1, 10, or 100), we generated 81 samples and 3,240 pairwise comparisons. For each pair, we computed four abundance-weighted dissimilarity metrics: Bray-Curtis with relative abundance (), Bray-Curtis with absolute abundance (), Weighted UniFrac with relative abundance (), and Weighted UniFrac with absolute abundance (). These comparisons help to illustrate how each axis of ecological difference—abundance, composition, and phylogeny—is expressed by the different metrics.

does not consistently yield higher or lower distances but instead varies depending on how abundance and phylogeny intersect (Fig. 1B). In the improbable scenario that all branch lengths are equal, is always less than or equal to (Fig. S1). These comparisons emphasize that once branch lengths differ, incorporating phylogeny and absolute abundance alters the structure of the distance space. The direction and magnitude of that change depend on which branches carry the abundance shifts.

To better understand how these metrics diverge, we examined individual sample pairs (Fig. 1C). Scenario 1 (gold star) illustrates the classic advantage of UniFrac: ASV\_1 and ASV\_2 are phylogenetically close, so and discern greater similarity between samples than and , which ignore phylogenetic structure. Scenario 2 (red star) highlights a limitation of relative metrics: two samples with identical relative composition but a 100-fold difference in biomass appear identical to and , but not to their absolute counterparts. In Scenario 3 (green star), incorporating absolute abundance decreases dissimilarity. and are lower than their relative counterparts because half the community is identical in absolute abundance, even though their proportions differ. In contrast, Scenario 4 (blue star) shows that can exceed when abundance differences fall on long branches, magnifying evolutionary distance and amplifying phylogenetic dissimilarity.

*Figure 1. Simulated communities reveal how absolute abundance affects phylogenetic and non-phylogenetic β-diversity measures.* (A) We constructed a simple four-ASV community with a known phylogeny and generated all permutations of each ASV having an absolute abundance of 1, 10, or 100, resulting in 81 unique communities and 3,240 pairwise comparisons. (B) Distributions of pairwise differences between weighted UniFrac using absolute abundance () and three other metrics: Bray-Curtis using relative abundance (), weighted UniFrac using relative abundance (), and Bray-Curtis using absolute abundance (). (C) Illustrative sample pairs demonstrate how absolute abundances and phylogenetic structure interact to increase or decrease dissimilarity across metrics. Colored stars indicate where each scenario falls within the distributions shown in panel B. Actual values for each metric are displayed beneath each scenario.

Across all 3,240 pairwise comparisons, is usually smaller than *BCA* and is more strongly correlated with (Pearsons = 0.82, < 0.0001) than with ( = 0.41) and ( = 0.55), reflecting the effect illustrated in Scenario 1. However, exceptions like Scenario 4 show that can also exceed when abundance differences occur on long branches. These scenarios demonstrate that integrates variation along three ecologically relevant axes: composition, phylogenetic relatedness, and microbial load, rather than isolating any single dimension. Because a given value can reflect multiple drivers of community change, interpreting it requires downstream analyses to disentangle the relative contributions abundance, phylogeny, and composition. To evaluate how this plays out in real systems, we next re-analyzed A graph of different types of data

AI-generated content may be incorrect.four previously published datasets spanning diverse microbial environments.

*Figure 2. Absolute UniFrac (UA) compared with other β-diversity metrics across four real microbial datasets.* Each panel shows pairwise sample distances for *UA* (x-axis) against another metric (y-axis): Bray-Curtis using relative abundance (*BCR*, first column), weighted UniFrac using relative abundances (*UR*, second column), and Bray-curtis using absolute abundances (*BCA*, third column). Contours indicate the density of pairwise comparisons (n shown for each dataset, left), with darker shading corresponding to more observations. The dashed line marks the 1:1 relationship. Points above the line indicate cases where *UA* is smaller than the comparator metric, while points below the line indicate cases where *UA* is larger.

***Application of Absolute UniFrac to Four Real-World Microbiome Datasets***

To illustrate the sensitivity of to both variation in composition and absolute abundance, we re-analyzed four previously published datasets from diverse microbial systems. These included: (i) nuclear reactor cooling water sampled across three reactor cycle phases [4]; (ii) the mouse intestinal track sampled across gut regions and under two diets [3]; (iii) depth-stratified freshwater communities sampled across two months [10]; and (iv) peanut rhizosphere microbiomes under two crop rotation regimes (conventional rotation (CR) vs. sod-based rotation (SBR)), plant maturities, and irrigation treatments [11]. Absolute abundance was quantified by flow cytometry for the cooling water and freshwater datasets, droplet digital PCR for the mouse gut, and by qPCR for the rhizosphere dataset. Together these span a wide range of richness–from 215 ASVs in the cooling water to 24,000 ASVs in the soil–and microbial load–from as low as 3.99x105 cells/ml (cooling water) up to 2x1012 16S rRNA copies/gram (mouse gut). Additional details of the re-analysis workflow, including ASV generation and phylogenetic inference, are provided in the Supporting Methods. Collectively, these re-analyses highlight that integrates ecological realism by incorporating differences in both lineage identity and total biomass.

We first calculated four β-diversity metrics for all sample pairs in each dataset and compared them to *UA* (Fig. 2). The degree of concordance between *UA* and other metrics was highly context dependent*.* In the cooling water dataset, *UA* closely tracked all three alternatives, whereas in the remaining systems it diverged substantially. *UA* also spanned a similar or wider range of distances than the other metrics. For example, in the soil dataset *BCR* and *UR* occupied a narrow range relative to the broad separation observed under *UA.*

*UA* generally reported distances that were similar to or greater than *UR,* consistent with the simulations shown in Fig. 1B. This reflects the ability of *UA* to discern dissimilarity due to differences in microbial load, even when community composition is conserved. In contrast, *UA* yielded distances that were similar to or lower than *BCA,* again matching the simulated behavior in Fig 1B. In these cases, phylogenetic proximity between abundant, closely-related ASVs leads *UA* to register greater similarity than *BCA.*

Given these differences, we next quantified how well each metric discriminates among categorical sample groups. For each dataset, we ran PERMANOVAs to measure the proportion of variance (*R2*) and statistical power (*pseudo-F, p-*value) attributable to group structure, using groupings that were determined to be significant in the original publications. To evaluate how strongly absolute abundance contributed to this discrimination, we also calculated *GUA* across a range of values. The resulting *R2* values are displayed in Fig. 3A, with the corresponding *pseudo-F* statistics and *p*-values provided in Fig. S2.

As in Fig. 2, the performance of each metric was strongly context dependent (Fig. 3A). In the mouse gut and freshwater datasets, absolute-abundance-aware metrics (*BCA* and *GUA*) explained the greatest proportion of variance (*R2*), and *R2* generally increasing as increased. In contrast, relative metrics captured more variation in the cooling water dataset (again at higher , and all metrics explained comparably little variance in the soil dataset. Taken at face value, these trends might suggest that higher values typically improve group differentiation.

However, this comes with a major caveat:

*Figure 3. Discriminatory performance of UA and related metrics across four microbial systems.* (A) PERMANOVAs were used to quantify the percent variance (*R2*) explained by predefined categorical groups (shown in italics beneath each dataset name), with 1,000 permutations. Metrics were evaluated across five metrics and, where applicable, across eleven values (0-1 in 0.1 increments). For consistency with the original studies, only samples from Reactor cycle 1 were used for the cooling-water dataset, only stool samples for the mouse gut dataset, and only mature rhizosphere samples for the dataset. (B) Mantel correlation (R) between each distance metric and the pairwise differences in absolute abundance (cell counts or 16S copy number), illustrating the degree to which each metric is driven by biomass differences.

We recommend calibrating based on research goals, mitigating this effect by using across a range of values rather than relying on or a single value. Researchers should decide how strongly they want their dissimilarity metric to place on microbial load. One potential approach is to calculate correlations as demonstrated in Fig. 3B and select an prior to any ordinations or statistical testing. Again, absolute metrics are expected to correlate to differences in cell count, as we hope absolute abundance measures incorporate differences in cell counts, especially when microbial load is relevant to the hypotheses being tested. Correlations to cell count in *BCA,* an accepted approach in the literature, ranged from ~0.5 up to ~0.8. As a general recommendation from these analyses, we recommend values in an intermediate range from 0.25 up to 0.75.

***Computational and methodological considerations***

A collage of graphs and diagrams

AI-generated content may be incorrect. Applying *GUA* in practiceraises several considerations related to sequencing depth, richness, and computational cost. Both Bray-Curtis and Unifrac can be sensitive to sequencing depth because richness varies with read count [13–15]. Methods to address these concerns, including rarefaction, remain debated and are challenging to benchmark rigorously [16]). We do not evaluate the sensitivity of *UA* or *GUA* to different rarefaction strategies here, but we provide a workflow and code for how we incorporated rarefaction into our own analyses (Fig. S4 and available code). Briefly, we rarefied our ASV tables to an even sequencing depth across 100 iterations; transform each ASV to relative abundance; normalize the abundance of each ASV in each sample by multiplying the relative abundance by the absolute abundance of that sample; and average the distance measure across all iterations. This approach minimizes sequencing-depth biases while preserving abundance scaling for downstream -diversity analysis. Future work evaluating how *UA* responds to richness and sequencing depth, relative to *UR* and *BCA,* would further clarify its performance characteristics.

*Figure 4. GUA requires more computational time but remains resilient to quantification error.* (A) Runtime for *GUA* (GUnifrac), *UR* (FastUnifrac in the phyloseq) and *BCA* (vegan) was benchmarked across 50 iterations on a sub-sampled soil dataset (mature samples only) [11], using increasing ASV richness (50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000), with 10 samples and one value per run (unweighted Unifrac is also calculated by default). (B) Linear relationships between *GUA* computation time and ASV richness. (C-D) Sensitvity of *GUA* ( = 1 and 0.2) and *BCA* to measurement error was evaluated by adding random variation (±1% to ±50%; Supporting Methods) to 16S Copy number estimates in the mouse gut dataset [3]. For each error level, 50 replicate matrices were generated and compared to the original values. Panels reflect the (C) mean difference and (D) max difference between the error-added metrics compared to the originals. Error bars represent the standard deviation of the average mean and max difference across 50 iterations.

*GUA* is slower to compute than both *BCA* and *UR* because it must traverse the phylogenetic tree to calculate branch length for each calculation. Computational time of *GUA* increases linearly with ASV number and is (Fig. 4A-B). The number of samples and values, however, have relatively little effect on runtime (Fig. S5). For datasets of ~10,000 ASVs, computing with a single CPU is still reasonable (~6 minutes), but repeated rarefaction increases runtime substantially because branch lengths are redundantly calculated with each iteration. Allowing branch-length objects to be cached or incorporated directly into the GUnifrac package workflow would considerably improve computational efficiency.

We also evaluated the sensitivity of *GUA* to measurement error in absolute abundance due to uncertainty arising from the quantification of cell number of 16S copy number. We added random error to the 16S copy number measurements from the mouse gut dataset, limiting our analyses to the stool samples where copy numbers varied by an order of magnitude. Across 50 iterations, each copy number could randomly vary by a given percentage of error in either direction. We re-calculated -diversity (*BCA* and *GUA* at = 1 and = 0.2) and compared these measurements to the original dataset.

All three measurements were resilient to added error (Fig. 4C). At an = 1, for every 1% of random variation we added to absolute quantification *GUA* differed from the original measurement by only 0.0022; even at ±50% added noise, *GUA* only differed (on average) by 0.1. And while *GUA* ( = 1) was more sensitive to error than *BCA,* at = 0.2 it was *less* sensitive, again highlighting the potential benefits of moderate values. The max deviation from true that added error could inflict on a given metric was also proportional (and always less) than the magnitude of the error itself (Fig. 4D). Put simply, if one adds 10% of random error to a measurement, they can expect *GUA* to change, at max, by 0.1; in general the deviation from true will be much less. A more rigorous approach to assessing error propagation within these metrics, including mathematical proofs of the relationships estimated above, is outside the scope of this paper but would be helpful.

**Ecological Interpretation and conceptual significance**

Absolute Unifrac reframes the interpretation of *β*-diversity by making biomass an explicit ecological axis rather than an unmeasured or normalized-away quantity. Conventional UniFrac capture composition and shared evolutionary history, but implicitly invites interpretation as if it also encodes differences on microbial load. By incorporating absolute abundance directly, Absolute UniFrac helps to resolve this mismatch and restores alignment between ecological hypotheses and the quantities represented in the metric. In this view, β-diversity becomes a three-axis ecological measure, incorporating composition, phylogeny and biomass, rather than a two-axis approximation.

There are many cases where the incorporation of absolute abundance allows microbial ecologists to assess more realistic, ecologically-relevant differences in microbial communities, especially when microbial load is mechanistically central. Outside of the datasets re-analyzed here [3, 4, 10, 11], the temporal development of the infant gut microbiome involves both a rise in absolute abundance and compositional changes [6]; bacteriophage predation in wastewater bioreactors can be understood only when microbial load is considered [17]; and antibiotic-driven declines in specific swine gut taxa were missed using relative abundance approaches [18]. As -diversity metrics (and UniFrac specifically) remain central to microbial ecology, these findings highlight how interpretation changes once biomass is incorporated. Broader adoption of absolute abundance profiling will also depend on data availability. Few studies currently make absolute quantification data publicly accessible, underscoring the need to to deposit absolute measurements alongside sequencing reads for reproducibility, ideally as metadata within SRA submissions.

While demonstrated here with 16S rRNA data, the approach should extend to other marker genes or (meta)genomic features provided absolute abundance estimates are available. In this sense, offers not only a more ecologically grounded view of lineage turnover by jointly reflecting variation in biomass and phylogenetic structure.

No single metric (or in *GUA*), however, is universally “best”. Each β-diversity metric emphasizes a different dimension of community change. Researchers should therefore select metrics based on the ecological quantity that is hypothesized to matter most. Here, we demonstrate notthat *GUA* outperforms other measures, but that it faithfully incorporates the three axes of variation it was designed to incorporate: composition, phylogenetic similarity, and absolute abundance (Fig. 1 and Fig. 2). As with any -diversity analysis, interpretation requires matching the metric to the ecological question and exploring sensitivity across where applicaropriate [19]). By providing demonstrations and code for the application and interpretation of *UA/GUA,* we hope to encourage the use of these metrics as a tool of microbial ecology.

**Box 1: Interpreting absolute UniFrac in hypothesis-driven analyses**

Absolute UniFrac is most informative when variation in microbial load is expected to carry ecological meaning rather than being a nuisance variable. The choice of α determines how strongly abundance differences influence the metric, and should therefore be selected based on the hypothesis, not by convention. In settings where biomass is central to the mechanism under study, higher α values appropriately foreground that signal, whereas in cases where load variation is incidental or confounding, lower α values maintain interpretability. Framing α as a hypothesis-driven choice repositions β-diversity from a default normalization step to an explicit ecological decision.

***Decision framework for selecting alpha***: Researchers can also use this framework in reverse: if exploratory analyses show that α strongly alters group separation, that sensitivity itself may indicate that biomass is an untested but relevant ecological dimension of community change.

Hypothesis Type Beta metric Appropriate alpha Ecological Example

Biomass is mechanistically central (high alpha, maybe 0.5-1), algal or pathogen blooms?

Biomass matters but interact with other axes (moderate alpha 0.2-0.5), infant gut?, disturbance recovery, early succession, community shifts due to nutrient fluxes

Biomass differences are incidental or confounding (low alpha 0-0.2), spatial niche portioning, host filtering independent of growth, compositional shifts but constant microbial load.

**Final synthesis**

By explicitly incorporating absolute abundance, Absolute UniFrac shifts phylogenetic β-diversity from a two-axis approximation to a three-axis ecological measure. This reframing connects the metric to the underlying biological questions that motivate most microbiome studies, rather than to default normalization conventions. As methods for quantifying microbial load continue to expand, the ability to interpret β-diversity in a biomass-aware framework will become increasingly important for distinguishing true ecological turnover from proportional change alone. In this way, Absolute UniFrac is not simply an alternative distance metric but a tool for aligning statistical representation with ecological mechanism.

**References**

1. Gloor GB et al. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* 2017;**8**:2224. https://doi.org/10.3389/fmicb.2017.02224

2. Vandeputte D et al. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 2016;**65**:57–62. https://doi.org/10.1136/gutjnl-2015-309618

3. Barlow JT, Bogatyrev SR, Ismagilov RF. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. *Nat Commun* 2020;**11**:2590. https://doi.org/10.1038/s41467-020-16224-6

4. Props R et al. Absolute quantification of microbial taxon abundances. *ISME J* 2017;**11**:584–587. https://doi.org/10.1038/ismej.2016.117

5. Wang X et al. Current Applications of Absolute Bacterial Quantification in Microbiome Studies and Decision-Making Regarding Different Biological Questions. *Microorganisms* 2021;**9**:1797. https://doi.org/10.3390/microorganisms9091797

6. Rao C et al. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature* 2021;**591**:633–638. https://doi.org/10.1038/s41586-021-03241-8

7. Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 2005;**71**:8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

8. Lozupone CA et al. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Applied and Environmental Microbiology* 2007;**73**:1576–1585. https://doi.org/10.1128/AEM.01996-06

9. Chen J et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* 2012;**28**:2106–2113. https://doi.org/10.1093/bioinformatics/bts342

10. Pendleton A, Wells M, Schmidt ML. Upwelling periodically disturbs the ecological assembly of microbial communities in the Laurentian Great Lakes. 2025. bioRxiv, 2025. , 2025.01.17.633667

11. Zhang K et al. Absolute microbiome profiling highlights the links among microbial stability, soil health, and crop productivity under long-term sod-based rotation. *Biol Fertil Soils* 2022;**58**:883–901. https://doi.org/10.1007/s00374-022-01675-4

12. Morton JT et al. Uncovering the Horseshoe Effect in Microbial Analyses. *mSystems* 2017;**2**:10.1128/msystems.00166-16. https://doi.org/10.1128/msystems.00166-16

13. Schloss PD. Evaluating different approaches that test whether microbial communities have the same structure. *The ISME Journal* 2008;**2**:265–275. https://doi.org/10.1038/ismej.2008.5

14. Fukuyama J et al. Comparisons of distance methods for combining covariates and abundances in microbiome studies. Biocomputing 2012. WORLD SCIENTIFIC, 2011, 213–224.

15. McMurdie PJ, Holmes S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology* 2014;**10**:e1003531. https://doi.org/10.1371/journal.pcbi.1003531

16. Schloss PD. Waste not, want not: revisiting the analysis that called into question the practice of rarefaction. *mSphere* 2023;**9**:e00355-23. https://doi.org/10.1128/msphere.00355-23

17. Shapiro OH, Kushmaro A, Brenner A. Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. *The ISME Journal* 2010;**4**:327–336. https://doi.org/10.1038/ismej.2009.118

18. Wagner S et al. Absolute abundance calculation enhances the significance of microbiome data in antibiotic treatment studies. *Front Microbiol* 2025;**16**. https://doi.org/10.3389/fmicb.2025.1481197

19. Kers JG, Saccenti E. The Power of Microbiome Studies: Some Considerations on Which Alpha and Beta Metrics to Use and How to Report Results. *Front Microbiol* 2022;**12**. https://doi.org/10.3389/fmicb.2021.796025