**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**

Microbial ecologists routinely use -diversity metrics to compare communities, yet these metrics vary in the ecological dimensions they capture. Popular for incorporating phylogenetic relationships, UniFrac distances default to relative abundance, omitting important variation in microbial load. As methods for estimating absolute abundance gain traction, incorporating this information into -diversity analyses becomes essential. Here, we present *Absolute UniFrac* (), a variant of Weighted UniFrac that uses absolute abundances. Through simulations and a reanalysis of four 16S metabarcoding datasets (from a nuclear reactor cooling tank, the mouse gut, a freshwater lake, and the peanut rhizospere), we demonstrate the nuanced ways Absolute UniFrac incorporates microbial load and phylogenetic relationships in comparison to other abundance-aware metrics. While this can improve statistical power to detect ecological shifts, we also find Absolute Unifrac can be strongly correlated to differences in cell abundances alone. We therefore recommend a generalized form () with a tunable parameter to balance sensitivity and interpretability, and recommend testing the resilience of a given conclusion across multiple parameters and -diversity metrics. In terms of the computational application of *GUA* we demonstrate that *GUA* is significantly slower than other -diversity metrics. Finally, we show *GUA* is comparably insensitive to noise within the quantification of absolute abundance compared to Bray-Curtis dissimilarities, especially at lower values of

**Main Text**

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].

To overcome compositional constraints, researchers increasingly use flow cytometry, qPCR, and genomic spike-ins to quantify microbial load [4, 5]. These tools improve detection of functionally relevant taxa and mitigate the compositional constraints imposed by sequencing [1, 2]. Most studies using absolute data have used Bray-Curtis dissimilarity, which does not necessarily expect normalization to proportions [4, 6]. But in the field of microbial ecology, UniFrac distances remain popular when working with relative abundance data. Here, we present *Absolute UniFrac*, a direct extension of Weighted UniFrac that incorporates total abundance, and evaluate its impact across simulated and real-world datasets.

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 we weight the length of each branch, , 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.

Because is most sensitive to changes in abundant lineages, it 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 :

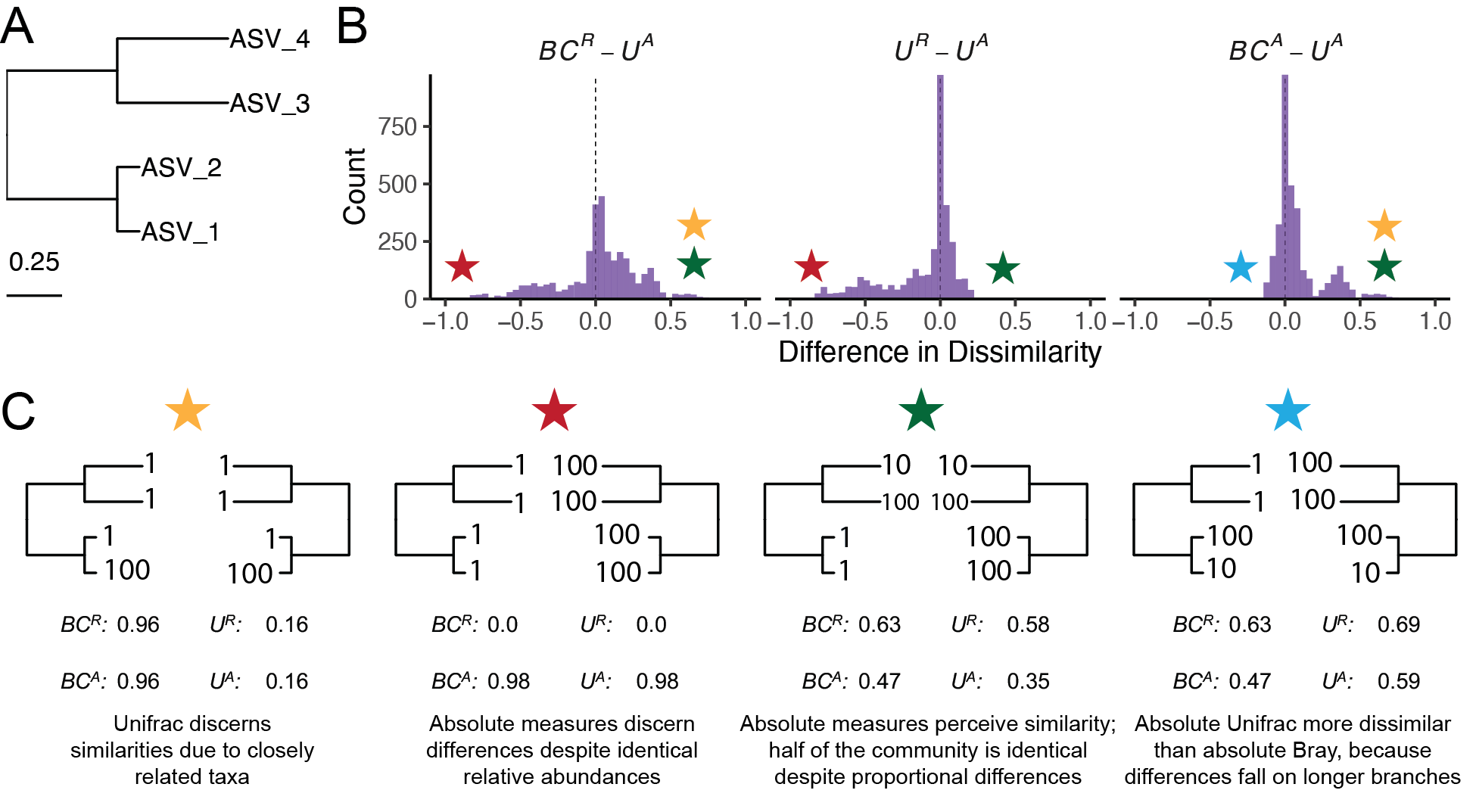
where ranges from 0 (close to Unweighted UniFrac) up to 1 (identical to , above). However, if one wishes to use absolute abundances, both and can be derived without normalizing to proportions:

Where and stands for the absolute counts of species descended from branch in community and , respectively. We refer to these distances as *Absolute Unifrac* and *Generalized Absolute Unifrac* ( and ). As can be seen, this replacement is mathematically trivial, but we were unable to find any examples in the literature where absolute abundances are discussed in relation to Unifrac distances, either conceptually or in application. Absolute abundances add another axis along which can vary; besides differences in community composition and phylogenetic similarity, also considers absolute differences.

It is a nontrivial task to consider how behaves in comparison to other metrics of using -diversity, especially in complex microbiomes with many samples. To first see how compares to other metrics, we constructed a simple example: a simulated community of four ASVs arranged in a simple phylogeny (Fig. 1A). By varying the absolute abundance of each ASV (1, 10, or 100), we generated 81 samples and 3,240 pairwise comparisons. For each pair, we computed four dissimilarity metrics: Bray-Curtis with relative abundance (), Bray-Curtis with absolute abundance (), Weighted UniFrac with relative abundance (), and Weighted UniFrac with absolute abundance ().

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 incorporating phylogeny and absolute abundance reshapes distance estimates in nontrivial ways.

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 terms, despite proportional differences. In contrast, Scenario 4 (blue star) shows that can increase dissimilarity relative to when abundance differences occur on long branches, 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) Examples of illustrative comparisons which reflect how absolute abundances and phylogenetic similarity can drive differences across dissimilarity 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 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 yield larger distances than when abundance differences occur on long branches. These scenarios demonstrate that can integrate changes along multiple, ecologically relevant axes (including differences in composition, phylogenetic similarity, and microbial load). However, because a given distance is the result of multiple drivers of variation between communities, its interpretation likely requires additional data analyses – are differences being driven primarily by changes in absolute abundance, or composition?

A graph of different types of data

AI-generated content may be incorrect.*Figure 2. Differences between UA and other metrics across real microbial communities.* Four 16S datasets with absolute abundance quantification were reanalyzed to produce ASVs and four dissimilarity metrics were calculated. *UA* is always on the x-axis, while another metric is on the y-axis: *BCR* on the first column, *UR* in the second, and *BCA* in the third. Contours in each panel reflect the density of points within that space (total n given below each dataset name), with darker colors corresponding to more observations within that area. The 1:1 line, where each two metrics were equal, is drawn as a dashed line. When density falls above the 1:1 line, *UA* tended to lower dissimilarity than the other metric.

To illustrate the sensitivity of to variation in composition and absolute abundance, we re-analyzed four previously published datasets from diverse microbial systems. These datasets included: nuclear reactor cooling water across three reactor cycle phases [4]; mouse gut samples across multiple intestinal locations and between two diets [3]; freshwater samples from Lake Ontario taken at two months and multiple depths [10]; and peanut rhizophere samples between two crop rotation schemes (conventional rotation (CR) vs. sod-based rotation (SBR), plant maturities, and irrigation schemes [11]. In the cooling water and freshwater datasets, absolute abundance was calculated via flow cytometry; in the mouse gut via droplet dPCR; and in the soil datasets, qPCR. These datasets also represent a wide range of richness–ranging from 215 total ASVs in the cooling water to 24,000 total ASVs in the soil–and microbial load–from as low as 399,120 cells/ml (cooling water) up to 2x1012 16S copies/gram (mouse gut). Additional details of the re-analysis, including ASV generation and phylogenetics, can be found in the Supporting Methods.

We first calculated four β-diversity metrics for all sample pairs in each dataset and compared them to *UA* (Fig. 2). Agreement between a given metric and UA is highly context dependent*.* In the cooling water, *UA* was in close agreement with all three other metrics, while it deviated broadly in other datasets. *UA* also measured distance values across a similar or wider range compared to other metrics; for example, notice how in the soil dataset, *BCR* and *UR* have narrow ranges compared to the broad range of *UA.*

*UA* measured either similar or greater distance than *UR,* in accordance with the simulations shown in Fig. 1B. We attribute this to the ability of *UA* to discern differences due to differences in microbial load, even when community composition is similar. In contrast, *UA* measured either similar or less distance than *BCA,* again in agreement with simulations shown in Fig 1B. Here, phylogenetic similarity between abundant, closely-related ASVs may allow *UA* to perceive greater similarity than *BCA.*

Given the broad differences in β-diversity measures within a given dataset, we next sought to quantify the ability of a given metric to separate microbial samples across treatments or categorical groups. We ran PERMANOVAs to assess the amount of variance and statistical power a given metric could discern between groups which were determined to be significantly different in their original publications. To assess the overall importance of absolute abundance in driving this discrimination, we also calculated *GUA* across a range of values The amount of variance (*R2*) from these PERMANOVA tests is displayed in Fig. 3A, while the *pseudo-F* statistic and *p*-values are provided in Fig. S2.

As in Fig. 2, the performance of a given metric is highly context dependent (Fig. 3A). In the mouse gut and freshwater datasets, absolute measures like *BCA* and *GUA* were able to explain the greatest amount of variance (*R2*); *R2* also increased as increased. In contrast, relative measures captured greater variance in the cooling water dataset (though again at higher , and all metrics captured a similar (and low) amount of variance in the soil dataset. This would initially lead to the conclusion that higher values are always better at differentiating groups.

However, this recommendation comes with a major caveat:

*Figure 3. The ability of UA to discriminate between groups or treatments.* (A) PERMANOVAs were run testing the significance of two-three category groups from each dataset (provided in italics beneath data names) with 1,000 iterations. Results indicate the percent variance explained (*R2*) for that variable.We assessed these metrics across five metrics and, where applicable, across eleven values (from 0 to 1 in 0.1 steps). Note that in the cooling reactor, only samples from Reactor cycle 1 were used; in the mouse gut, only stool samples were used, and in the soil, only mature samples were used. (B) Correlation between each given distance metric and the absolute difference in cell counts or 16S copy number between samples.

We urge careful calibration of based on research goals, thereby mitigating this effect by using across a range of rather than . Researchers should consider how much emphasis 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.

Finally, we sought to quantify some of the practical concerns researchers may have when applying *GUA* to their own data. Both Bray-Curtis and Unifrac measures can be sensitive to differences in richness as a result of uneven sequencing depth [13–15]. Methods to address these concerns, including rarefaction, are highly debated in the literature, and the simulations to test the impact of rarefaction decisions are complex (e.g. [16]). We do not explore the specific sensitivity of *UA/GUA* to rarefaction decisions here, but do provide a practical framework and code for how we incorporated rarefaction into our own analyses (Fig. S4 and available code). Put simply, we rarefy our ASV tables to an even sequencing depth across many iterations (here, 100); transform each ASV to relative abundance; normalize the abundance of each ASV in each sample by multiple the relative abundance by the absolute abundance of that sample; and average the distance measure across all iterations. Future work which assesses the sensitivity of *UA* to A collage of graphs and diagrams

AI-generated content may be incorrect.richness and sequencing depth, compared to *UR* and *BCA,* would be useful to the field.

*Figure 4. Performance of GUA in terms of computational time and resilience to quantification error.* (A) The computational time to calculate *GUA* (from the GUnifrac package), *UR* (implemented as FastUnifrac in the phyloseq package) and *BCA* (from the vegan package) was measured across 50 iterations on the sub-sampled soil dataset (only mature samples) [11] across a range of ASV numbers (50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000), holding a constant of 10 samples and 1 requested value (though unweighted Unifrac is also calculated by default). (B) Linear relationships between *GUA* computation time and ASV number. (C-D) Using only the stool samples from the mouse gut dataset [3], random error (ranging from ±1% variation up to ±50% variation) was added to the 16S copy number of each sample (additional details in Supporting Methods). Across 50 iterations, the difference between *BCA, GUA* ( =1) and *GUA* ( =0.2) on the error-added dataset was compared to the original calculation of that metric for each sample comparison. Points reflect the (C) mean difference and (D) max difference between the error-added metrics compared to the original calculation. Error bars represent the standard deviation of the average mean and max difference across 50 iterations.

*GUA* is slower to compute than other metrics, including *BCA* and *UR* (Fig. 4A). On average, *GUA* is two-four orders of magnitude slower than these metrics. The computation time of *GUA* is driven primarily, and linearly, by the number of ASVs in the dataset, as traversing the tree to calculate branch length is computationally expensive. In contrast, the number of samples or number of values to calculate weakly influences the computation time (Fig. S5). While computation time on a 10,000 ASV dataset with a single CPU is still reasonable (~6 minutes), when calculating *GUA* over many iterations of rarefaction the total computation time can become inconvenient, as branch lengths are redundantly calculated each time. If rarefaction could be incorporated into the GUnifrac package, or a branch-lengths object could be saved and re-used such that branch lengths are only calculated once, this would greatly improve computational efficiency.

It is also expected that estimation of -diversity metrics which rely on absolute abundance will be sensitive to errors or uncertainty arising from the quantification of cell number of 16S copy number. To assess the sensitivity of *GUA* and *BCA* to measurement error, 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.

There are many cases where the incorporation of absolute abundance allows microbial ecologists to assess more realistic, ecologically-relevant differences in microbial communities, especially in contexts where microbial load matters. Outside of the datasets re-analyzed here [3, 4, 10, 11]; the temporal development of the infant 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, we encourage researchers to adopt when applicable. In addition, our re-analysis efforts were limited by few papers sharing the absolute quantification data. The open sharing of absolute quantification data, ideally as metadata within SRA submissions, is just as essential for reproducibility as the sequencing files themselves.

While demonstrated here with 16S rRNA data, the approach should be generalizable to other marker genes or (meta)genomic features, provided absolute abundance estimates are available. In doing so, offers not only a more grounded picture of lineage differences but also sensitivity to both biomass variation and phylogenetic depth.

In reality, no one metric (or in *GUA*) is “better” than another – all metrics shown here calculate distance or dissimilarity along valid, but different, axes of variation in microbial communities. Researchers should select β-diversity measures to address their specific hypotheses on how two microbiomes may differ, and interpret the results of a β-diversity measure in terms of what information that measure incorporates. Here, we demonstrate **not** that *GUA* outperforms other measures, but that it does incorporate the axes of variation (composition, phylogenetic similarity, and absolute abundance) it was designed to incorporate (Fig. 1 and Fig. 2). As with any -diversity study, researchers should interpret results critically, explore sensitivity of results across different metrics, and justify their choice of metrics (and when applicable [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.

**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