Reviewer #1:

I thank the authors for their simple but useful generalization of weighted UniFrac distances from relative to absolute abundances. The key concepts are explained well and the properties and utility of the metrics are demonstrated and compared to others through toy examples as well as a real-world dataset. In general, it would be interesting to see comparison of metrics for more than one example, but I guess there are strict limitations for a brief communication.

> We thank the reviewer for their interest and encouragement. We have added substantially to the paper, including three additional data sets within our analysis, and an extended discussion (moving outside the constraints of the brief communication).

My only major comment is as follows:

To me, it seems trivial to replace relative abundance by absolute counts. Is this the first time this has been suggested or tried? Why isn't this already commonly used, given that suitable absolute abundance data is available?

> This is a useful comment, and reflects what was to us as a surprising gap in the literature. We added additional information (line 172) to emphasize that these metrics have (to our knowledge) not been used, and that while its derivation and application are simple, the resulting interpretation is nontrivial:
>
> "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 similarity, UA also captures divergence in microbial load. This makes interpretation of UA nontrivial, particularly in complex microbiomes."

In addition to this, I have some minor comments that should be addressed before publication:

99: It's informative to see how UA relates to the other metrics for specific examples, but I find it hard to draw the conclusion that it "integrates ecological realism" just from the numbers that are provided. Could you substantiate this claim more? I am also curious about how interpretable UA is compared to other metrics?

> We agree additional nuance is warranted. We've revised the text to reflect that UA is able to integrate multiple axes of important ecological variation, but that this integrative nature makes its interpretation more complex (line 176). This revision provides a natural transition to the new analyses presented in our study, as follows:
>
> Line 217: "These scenarios demonstrate that UA integrates variation along three ecologically relevant axes: composition, phylogenetic similarity, and microbial load, rather than isolating any single dimension. Because a given UA value can reflect multiple drivers of community change, interpreting it requires downstream analyses to disentangle the relative contributions of these three axes. To evaluate how this plays out in real systems we next reanalyzed four previously published datasets spanning diverse microbial environments.

Fig. 1B: Why aren't all the colored stars indicated in all the distributions in B?

We had originally added stars to the histograms sparingly, to draw attention to specific comparisons. However, stars have been added to all panels now.

Fig. 1C: Maybe I'm misunderstanding something, but aren't the first two trees in C (gold star) exactly the same, including the absolute abundances? In that case, shouldn't distances and dissimilarities be zero? Is there an error in the labels of one of the trees?

Thank you so much for catching this. We have corrected this mistake.

108: "UA yields greater or smaller dissimilarity than other metrics". UA and UR are the same in the first example.

We've clarified this sentence, so that it's not declarative of UA's relation to all other metrics, but reflects that these scenarios were chosen to illustrate specific cases where UA differs from at least one other metric in a conceptually important way. The new sentence reads:

Line 241: "Illustrative sample pairs demonstrate how absolute abundances and phylogenetic structure interact to increase or decrease dissimilarity across metrics.

142: "In this dataset, we recommend an intermediate of 0.5". This seems a bit arbitrary, and isn't the horseshoe still discernible for alpha = 0.5? Why do you recommend this value and how do you recommend choosing alpha in general?

Thank you for this question. We've updated our analyses to include a finer range of alphas (in steps of 0.1, rather than in 0.5), and included a new guidance for choosing an alpha, including Box 1 and lines 356-372, where we recommend tailoring alpha to better match the hypothesis being tested, exploring multiple levels of alpha, or choosing alpha *a priori* to modulate the impact of cell abundance on your diversity estimation using Mantel correlations.

158: "For example, the temporal development of the infant microbiome". It would have been interesting to see the different metrics applied to data for this example.

Unfortunately we were unable to access the necessary absolute abundance metadata to reanalyze that study. However, (as discussed elsewhere) we've incorporated analyses from three other datasets across a broad range of environments, richness, and evenness.

Reviewer #2

Major Comments:

As indicated by your own words on lines 81-83, "These comparisons emphasize that incorporating phylogeny and absolute abundance reshapes distance estimates in nontrivial ways.", there are novel properties associated with the newly proposed Absolute Unifrac and Generalized Absolute Unifrac metrics. These need to be understood through more than a toy four-ASV phylogeny and one freshwater 16S data set. It would be ideal to at least test:

1) data derived from multiple environments (e.g. soils, human gut) that represent a range of species diversity and abundances,

> Thank you for this thoughtful suggestion. We agree and have now incorporated three additional datasets from a nuclear cooling reactor, mouse guts, and soils, which range widely in richness and abundances, which we feel has greatly improved the manuscript. While analyzing even more datasets would be preferable, we struggled to find many studies that (1) used modern sequencing strategies appropriate for ASV generation via dada2, (2) quantified absolute abundance and provided that data publicly, and (3) provided sequencing data via the SRA with sufficient metadata to link samples to experimental treatments.

2) data derived from other marker genes (e.g. 18S, cytochrome oxidase, etc.),

> We agree that more analyses would be useful to the field. That said, it was already difficult to find studies with publicly available sequencing data, absolute abundance measurements, and high-quality metadata that were acceptable for reanalysis (see above). Studies incorporating absolute abundance are few and far between; studies with well-reported data are also unfortunately not as common as they should be, a point we now bring up in the manuscript itself. While one dataset we accessed (Zhang et al., peanut rhizosphere) did include fungal ITS sequences, we felt it wasn't additive to compare these results to the bacterial 16S results, without other ITS or 18S studies with which to compare it. That said, we found many examples where Weighted Unifrac was used on ITS or 18S datasets, reinforcing that the application of UniFrac distances is conceptually valid outside of 16S metabarcoding, specifically.

3) how the metric's validity is impacted by precision and error at the qPCR or flow-cytometry steps that could propagate into Absolute Unifrac distances,

> Great suggestion. We agree this is an important point. While a mathematical analysis of error propagation in UA vs. BCA is outside our expertise, we include simple simulations wherein error was added to absolute abundance measurements and GUA/BCA were re-calculated (Fig. 4C/D and corresponding results text, lines 485-501), demonstrating the general resilience of these methods to quantification error.

4) how this metric is impacted by rarefaction decisions.

On the last point re: rarefaction but also related to other points, the authors note on line 171 that "We also do not address…how sequencing depth influences richness estimates or whether rarefaction should be applied before calculating GUA.", but given that this is the paper introducing this new metric, and given that the metric directly relies on the abundance values which are the target of rarefaction efforts, it seems reasonable to expect guidance for readers on these steps. Likely, more testing is required to assess the impact of rarefaction on various data sets employing

GUA, but this key effort would lead to the creation of the guide readers require to apply this new computation tool.

> A thoughtful question! We agree that the effects of rarefaction are an important consideration, especially given that GUA directly incorporates abundance information. While a comprehensive evaluation of rarefaction strategies is beyond the scope of this paper, we have expanded the manuscript to provide conceptual guidance for readers. Specifically, we now include an overview of our rarefaction approach (Box 2) and publicly available implementation code (Github link). We hope this addition will help readers apply GUA in a consistent and transparent way and will motivate future studies explicitly evaluating how rarefaction influences this and related β-diversity metrics.

If such a guide could be produced to highlight the use cases for this new metric, it should be published. Ideally, everyone understands the conditions under which this metric outperforms other abundance-aware distances, of which there are many.

> This was a thought-provoking comment that influenced how we frame our results. We now emphasize that no single metric necessarily "outperforms" another, as each metric captures distinct yet valid axes of variation among microbial communities. In the revised discussion (lines 529-538) and Box 1, we highlight that the choice and interpretation of a metric should be guided by the specific hypotheses being tested and whether they concern compositional similarity, phylogenetic relatedness, or absolute abundance differences. Taken together, the manuscript demonstrates that UA integrates all three dimensions and provides concrete examples of how its results can be interpreted in ecological context.

Minor Comments:

The simulation data used for the phylogeny is far too simple and does not demonstrate ecological realism. Related to the comments about testing more data sets, the in-silico data can certainly be more complex and robust as a test data set.

> We appreciate this comment and agree that more complex simulations can be valuable for testing new metrics. However, the purpose of Fig. 1 is illustrative rather than analytical. The four-ASV community was intentionally designed as the simplest possible system to demonstrate, in a transparent way, how these metrics can respond to changes in composition, abundance, and phylogenetic relatedness. Increasing the complexity of this simulation would obscure these conceptual contrasts. We believe the inclusion of multiple new empirical datasets now addresses this concern more directly, as they capture the same patterns observed in Fig. 1 while providing the ecological realism the reviewer highlights.

The data and code available to produce the manuscript are all seemingly shared in a public Github. Kudos to the authors for this important step in providing a reproducible analysis.

> Thank you! We greatly appreciate this recognition, as ensuring full reproducibility is both deeply important and time-intensive. We have maintained this degree of reproducibility throughout all subsequent reanalyses as well.

If possible, for when data exists for a range of diverse data sets, it would be valuable to add any statements about computational efficiency relative to existing metrics (e.g. Bray-Curtis, Unifrac, Weighted Unifrac, etc.).

We appreciate this suggestion and have added new analyses to address it. Specifically, Fig. 4 and Fig. S4 now compare computational performances across metrics, showing that GUnifrac is slower than Bray-Curtis (vegan package) and Fast Unifrac (phyloseq package). We also provide recommendations for how future implementations of GUA, particularly across repeated rarefaction, could improve computational efficiency (line 482).

Figure S2 could indicate the sample labels, or at least a key for the labels, to aid in interpretability.

Thank you for this suggestion. As the manuscript now includes multiple datasets, we have restructured the figures and no longer include the original Fig. S2 and S3. The ordination based on UR (formerly Fig. S3) is now presented as a sub-panel in the new Fig. S3, which compares ordinations using UR and UA across all four datasets.