

**1 Interpretation and application of absolute abundance in**  
**2 Weighted UniFrac distance**

3 Augustus Pendleton<sup>1\*</sup> & Marian L. Schmidt<sup>1\*</sup>

4 <sup>1</sup>Department of Microbiology, Cornell University, 123 Wing Dr, Ithaca, NY 14850, USA

5 **Corresponding Authors:** Augustus Pendleton: arpt277@cornell.edu; Marian L.  
6 Schmidt: marschmi@cornell.edu

## Supplemental Figures

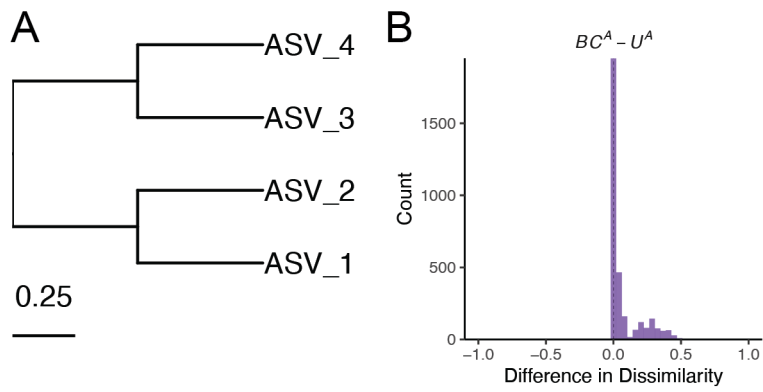
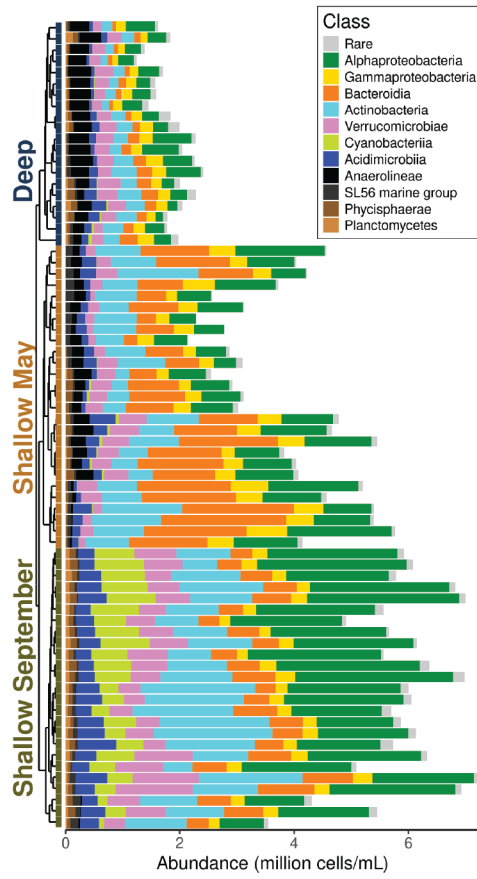
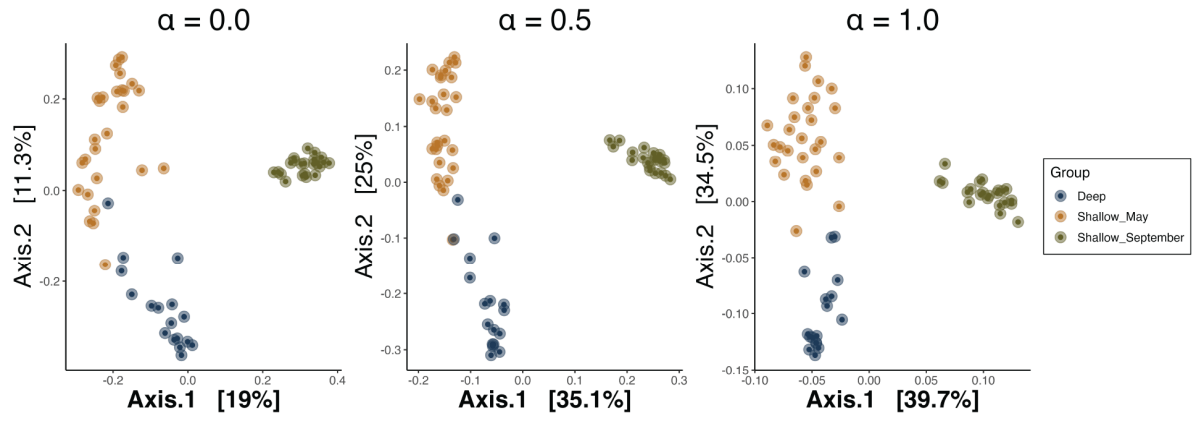


Figure S1.  $U^A$  is always less than  $BC^A$  when branch lengths are fully symmetrical. (A) Symmetrical tree used for simulations as opposed to non-symmetrical tree in Fig. 1A. (B) Distribution of differences between  $BC^A$  and  $U^A$ . As the differences are never negative,  $U^A$  is always less than or equal to  $BC^A$ .



12

13 *Figure S2. Taxonomic composition and absolute abundance of microbial communities from Lake Ontario.*  
 14 Each bar represents a sample taken from Lake Ontario. The height of the bar represents the absolute  
 15 abundance (cells/ml) in each sample, filled at the Class level. Samples are clustered via hierarchical  
 16 clustering (in this case, using UPGMA with  $GU^A$ ,  $\alpha = 0.05$ ), reflecting both changes in composition  
 17 (e.g. Cyanobacteriia in Shallow September, Anaerolineae in Deep) and absolute abundance (Deep samples  
 18 have the fewest cell counts, followed by Shallow May and Shallow September).



19

20 *Figure S3. Principal Coordinates Analysis of Lake Ontario samples using  $GUR$  across three  $\alpha$  values.*

Package/Software	Version	Citation
R	4.5.0	[1]
RStudio	2024.12.1+563	[2]
tidyverse	2.0.0	[3]
phyloseq	1.52.0	[4]
vegan	2.7-1	[5]
GUniFrac*	1.8.1	[6]
ggtree	3.16.0	[7]
patchwork	1.3.1	[8]
NatParksPalettes	0.2.0	[9]
ape	5.8-1	[10]
broom	1.0.8	[11]
corrr	0.4.4	[12]
renv	1.0.5	[13]

<sup>21</sup> *Table S1. Software and packages used in analysis.* Note that GUniFrac was modified  
<sup>22</sup> slightly to make incorporating absolute abundances more apparent; this version can be  
<sup>23</sup> installed via Github at <https://github.com/MarschmiLab/GUniFrac>.

## Supporting References

1. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2020.
2. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, PBC., 2020.
3. Wickham H. tidyverse: Easily install and load the tidyverse. 2023.
4. McMurdie PJ et al. phyloseq: Handling and analysis of high-throughput microbiome census data. 2022.
5. Oksanen J et al. vegan: Community ecology package. 2022.
6. 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>
7. Yu G et al. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 2017;**8**:28–36. <https://doi.org/10.1111/2041-210X.12628>
8. Pedersen TL. patchwork: The composer of plots. 2024.
9. Blake K. NatParksPalettes: Color palettes inspired by national parks. 2022.
10. Paradis E et al. ape: Analyses of phylogenetics and evolution. 2023.
11. Robinson D, Hayes A, Couch S. broom: Convert statistical objects into tidy tibbles. 2023.
12. Kuhn M, Jackson S, Cimentada J. corrr: Correlations in R. 2022.
13. Ushey K, Wickham H. renv: Project environments. 2024.