Table S1. Parameter values used in power study (2D circle-based simulation)

Community characteristic	Values for condition 1 (x_1)	Values for condition 2 (x_2)	Value of s
Centroid position (o)	$0 \sim 0.016$	$0 \sim -0.016$	0.03
Evenness (α)	$1 \sim 0.7$	$1 \sim 1.3$	0.3
Radius (r)	$0.15 \sim 0.132$	$0.15 \sim 0.168$	0.03
Scenario 4 Fold change (k)	1 (Fixed)	$1 \sim 2.5$	1
Scenario 5 Fold change (k)	1 (Fixed)	$1 \sim 2.8$	1
Scenario 6 Fold change (k)	1 (Fixed)	$1 \sim 6$	2

The values of x_1, x_2 are evenly spaced on a grid of 10. For Scenario 6, we decreased the abundances.

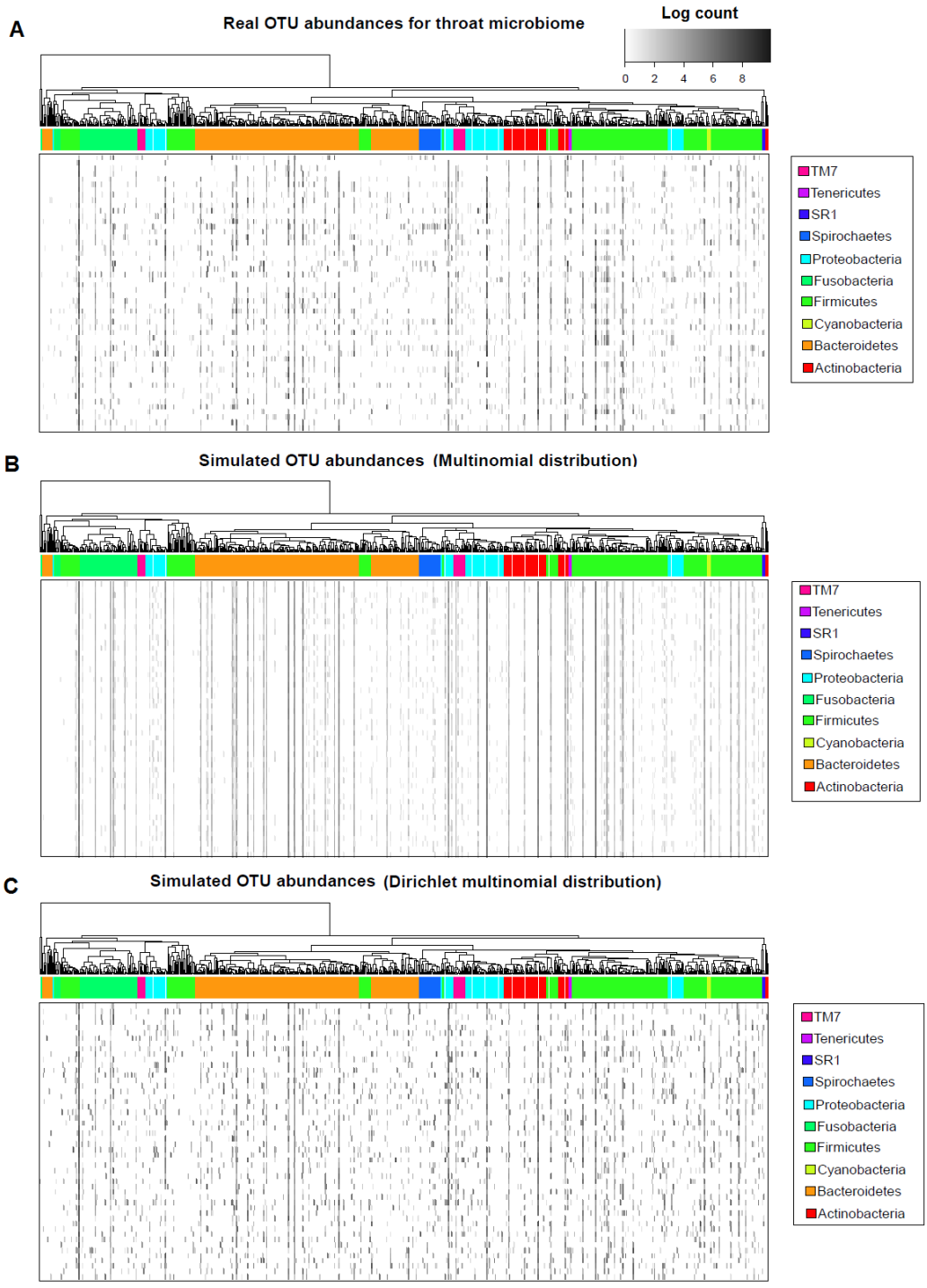

Figure S1. Comparison of Multinomial model and Dirichlet-Multinomial model for simulating OTU counts for a throat microbial community. (A) The heatmap shows the OTU abundance distribution from a real throat microbial community of 60 samples. Rows represent samples while columns correspond to OTUs. These OTUs are related by a phylogenetic tree colored by phyla. The gray scale indicates the level of abundance on a log scale with white meaning zero counts (see legend). (B) The OTU counts are generated by assuming a Multinomial model, where the parameters are estimated from (A). (C) The OTU counts are generated by assuming a Dirichlet multinomial (DM) model, where the parameters are estimated from (A). The DM models overdispersion better than Multinomial model.

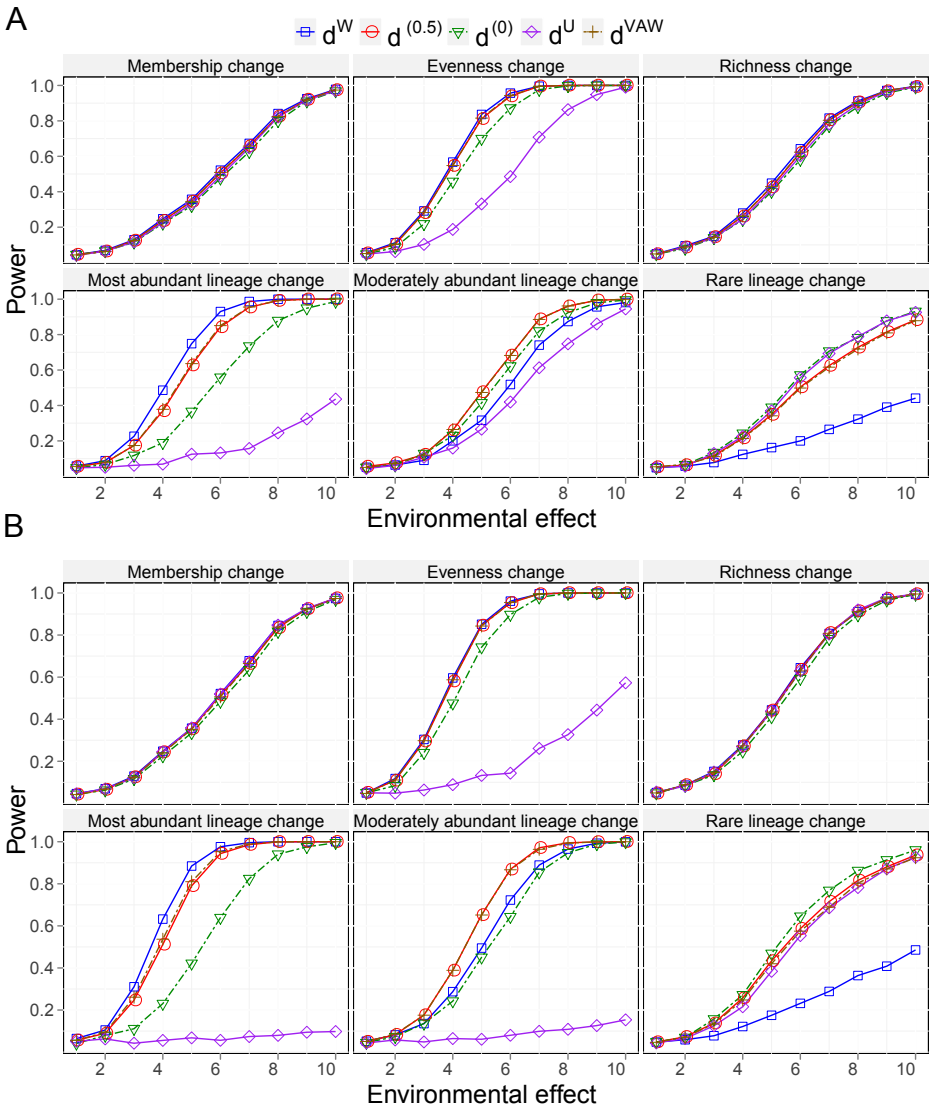

Figure S2. Power comparison of different UniFrac variants for detecting environmental effect using 2D circle based simulation and different bin sizes for OTU formation. Ten samples from each of the two environmental conditions are generated using 2D circle based simulation. A bin size of 0.01 (A) or 0.03 (B) is used in OTU formation. UniFrac distance matrices are constructed based on the simulated OTU abundances and NJ tree. PERMANOVA is used for testing hypotheses. d^W, $d^{(0.5)}$, $d^{(0)}$, d^U and d^{VAW} are compared and indicated by different colors. The specific community difference caused by different environmental conditions is indicated in the panel title. The power curves are created by varying the degree of environmental effect. The initial point of the power curve is the power when there is no environmental effect.

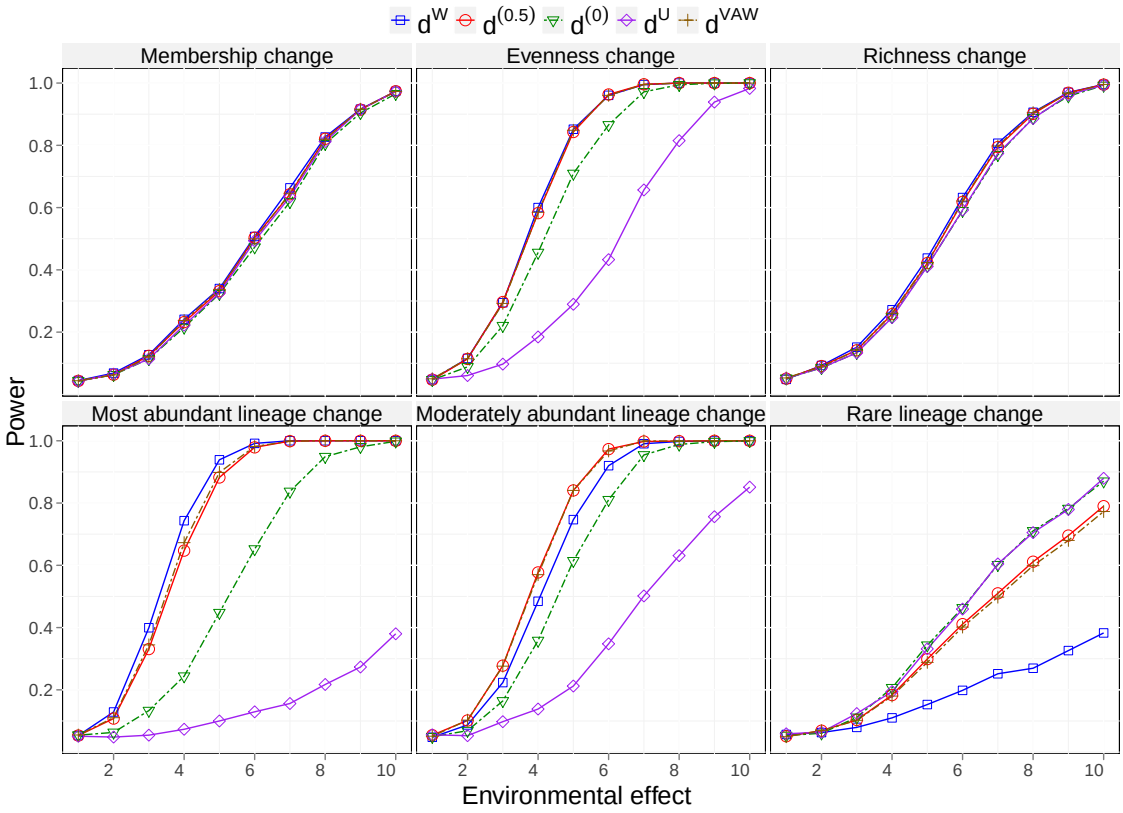

Figure S3. Power comparison of different UniFrac variants for detecting environmental effect using 2D circle based simulation and UPGMA tree. Ten samples from each of the two environmental conditions are generated using 2D circle based simulation. A bin size of 0.015 is used in OTU formation. UniFrac distance matrices are constructed based on the simulated OTU abundances and UPGMA tree. PERMANOVA is used for testing hypotheses. Four representative UniFrac variants d^W, $d^{(0.5)}$, $d^{(0)}$, d^U and d^{VAW} are compared and indicated by different colors. The specific community difference caused by different environmental conditions is indicated in the panel title. The power curves are created by varying the degree of environmental effects. The initial point of the power curve is the power when there is no environmental effect.

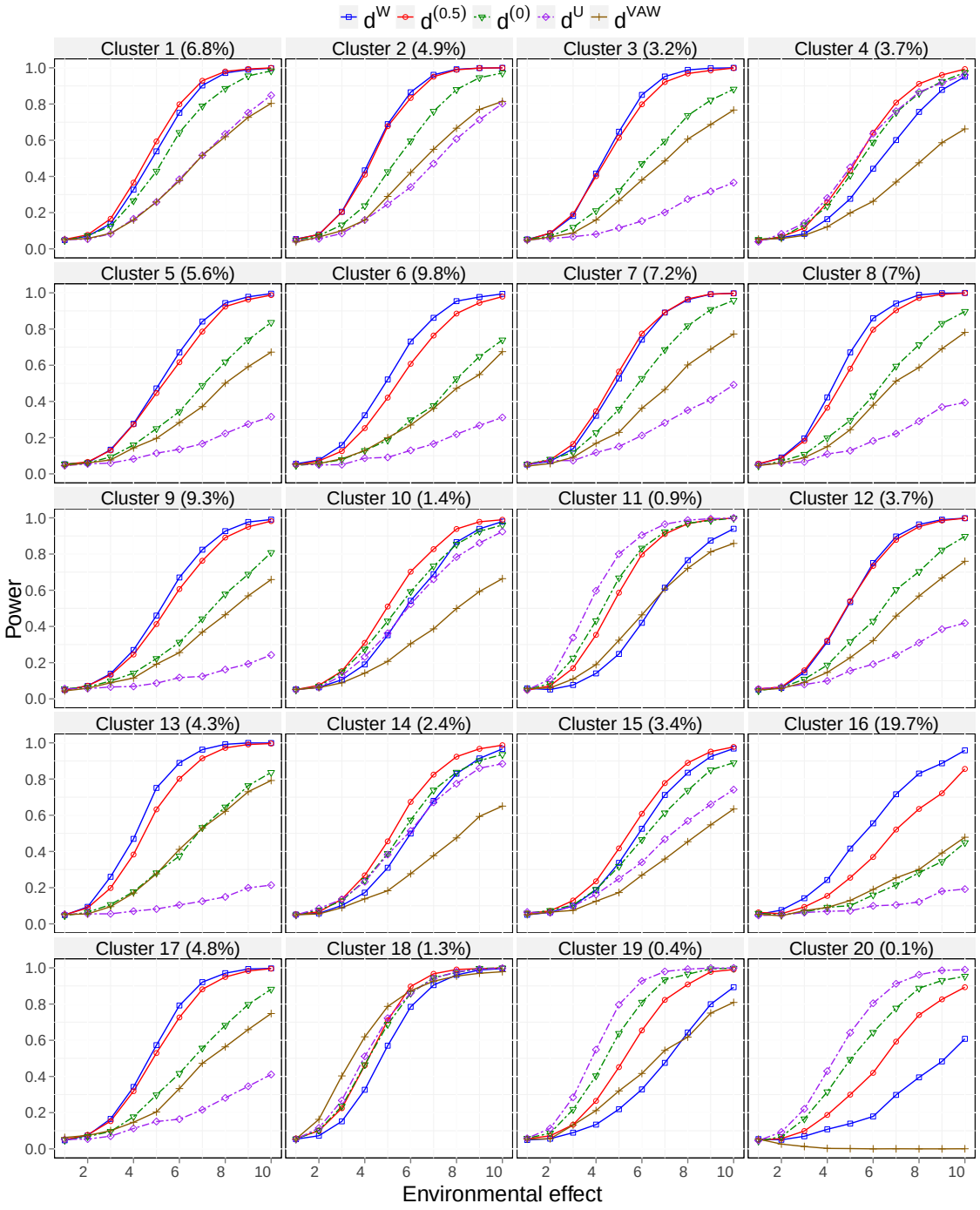

Figure S4. Power comparison of different UniFrac variants for detecting environmental effect using tree based simulation (all lineages). Ten samples from each of the two environmental conditions are generated using tree based simulation. UniFrac distance matrices are constructed based on the simulated OTU abundances and the phylogenetic tree. PERMANOVA is used for testing hypotheses. d^W, $d^{(0.5)}$, $d^{(0)}$, d^U and d^{VAW} are compared and indicated by different colors. The environmental factor affects a particular lineage (OTU cluster). The figure shows all the 20 lineages that are affected by environment. The lineage abundance is given in parentheses in the panel title. The initial point of the power curve is the power when there is no environmental effect.

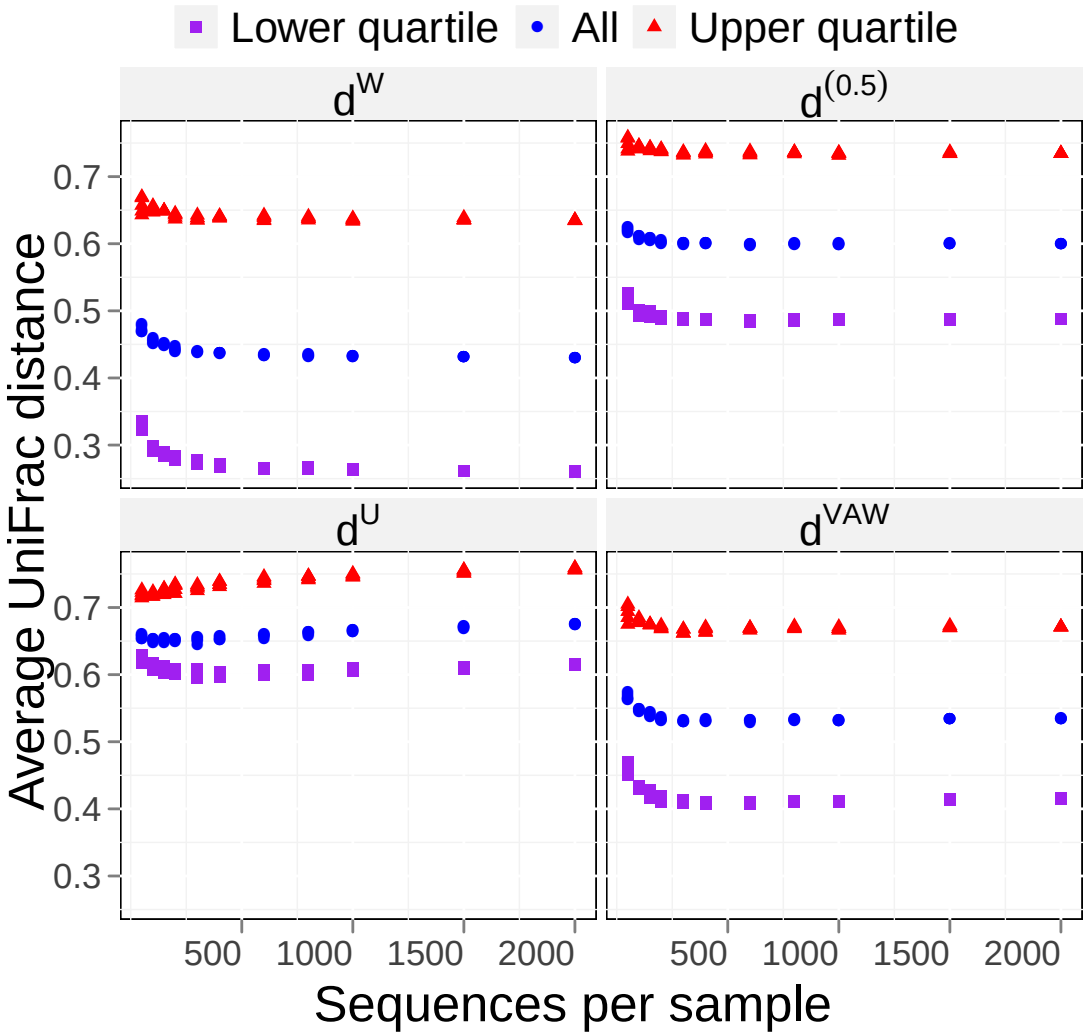

Figure S5. Rarefaction of data from a study of diet effect on the gut microbiome. This study produced about 1 million reads from the V12 region of 16S rRNA using pyrosequencing. The samples with less than 2408 sequences were first excluded (leaving 98 samples). For five replications, sequences from 98 samples were subsampled to different depth (between 50 to 2000). Pairwise distances were calculated for the four UniFrac variants (d^W, $d^{(0.5)}$, $d^{(U)}$ and d^{VAW}). To assess the the effects of community divergence on the sensitivity to sampling, the most similar and most different pairs of samples were identified from un-subsampled samples (2408 sequences) as those in the upper and lower quartile of UniFrac values calculated separately for all UniFrac variants. The points represent the average UniFrac value at each sampling depth for all pairs ('All') and the pairs that were in the upper and lower quartiles. Each point represents one replicate.

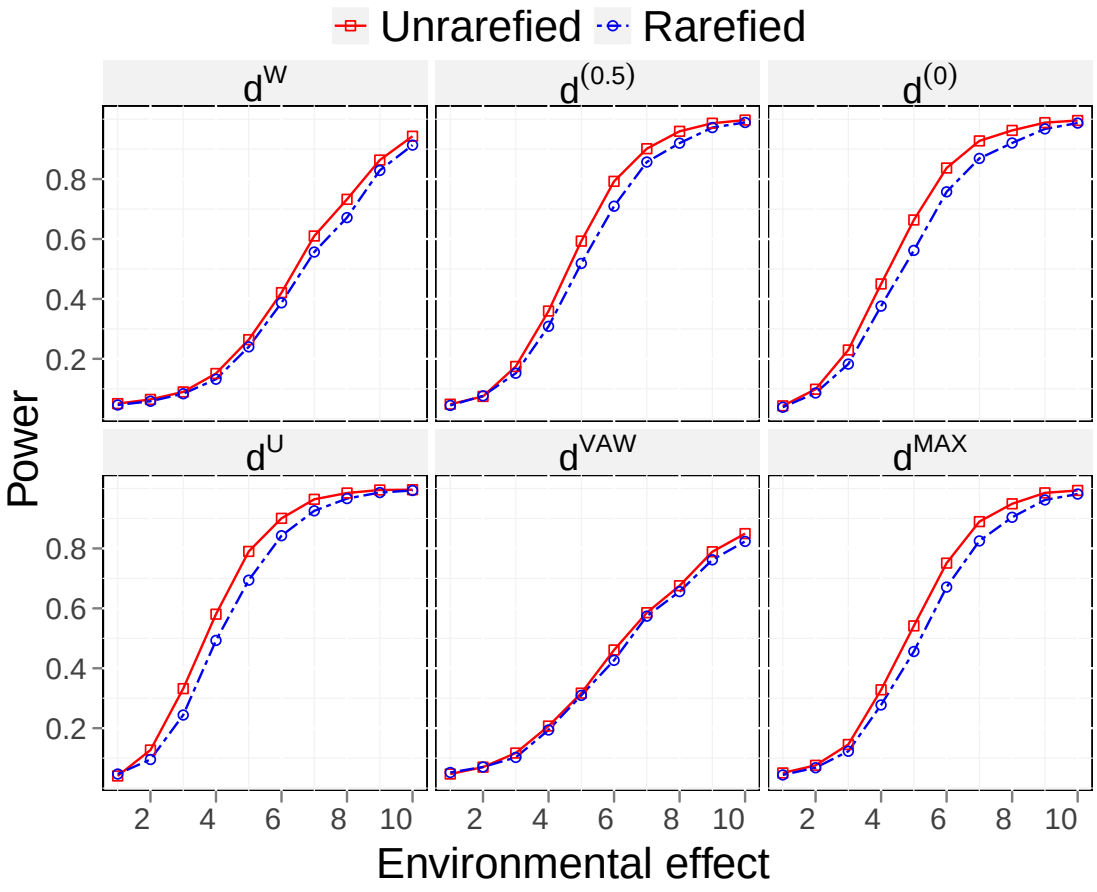

Figure S6. Effect of rarefaction on the power of testing the association of microbiome composition with covariates. The tree based simulation approach is used to investigate the effect of rarefaction on the power of PERMANOVA test. Two conditions are simulated with 10 samples under each condition. We let the environmental factor increases the abundance of OTU Cluster 11 of the tree for illustration purpose. Negative binomial (NB) model is used to generate the sampling depth for each sample. The parameters of the NB model is adjusted to have mean 1,000 and standard deviation (SD) of 300 so the sampling depth can range from 400 to 1600 (2SD) sequences per sample. We then compare the power of the test before and after rarefaction calculated over 2,000 replications. For the rarefaction case, we rarefied all the samples to the lowest sampling depth seen in the samples. We found that rarefaction decreases the power for all the UniFrac variants tested.