Supporting Information

Robeson et al. 10.1073/pnas.1012678108

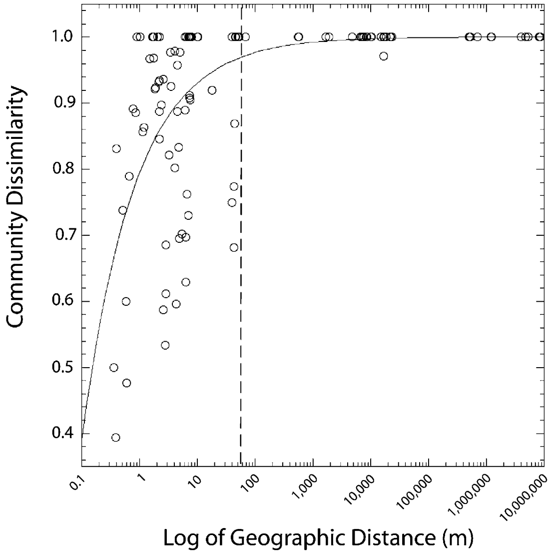

Fig. S1. Operational taxonomic unit (OTU)-based variogram using the Bray–Curtis metric (bdelloid abundance) as implemented in QIIME (1) plotted as community dissimilarity versus the log of geographic distance. A value of 1 indicates completely different communities, and a value of 0 indicates identical communities. The vertical dashed line indicates the autocorrelation range. Note the overestimation of community dissimilarity after the autocorrelation range.

1. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

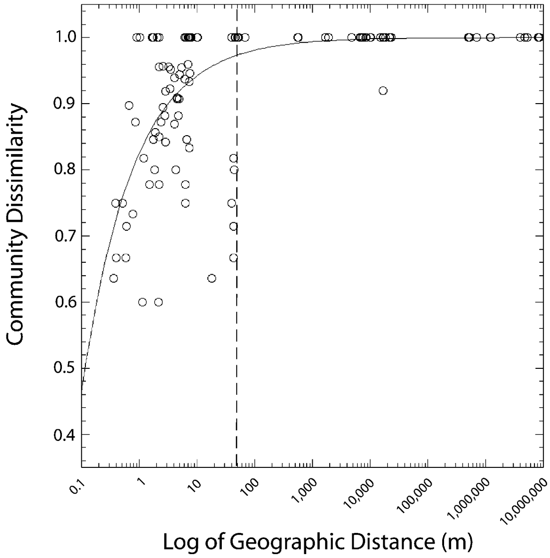

Fig. S2. OTU-based variogram using the Sorensen metric (bdelloid occurrence) as implemented in QIIME (1) and plotted as community dissimilarity versus the log of geographic distance. A value of 1 indicates completely different communities, and a value of 0 indicates identical communities. The vertical dashed line indicates the autocorrelation range. Note the overestimation of community dissimilarity after the autocorrelation range.

1. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

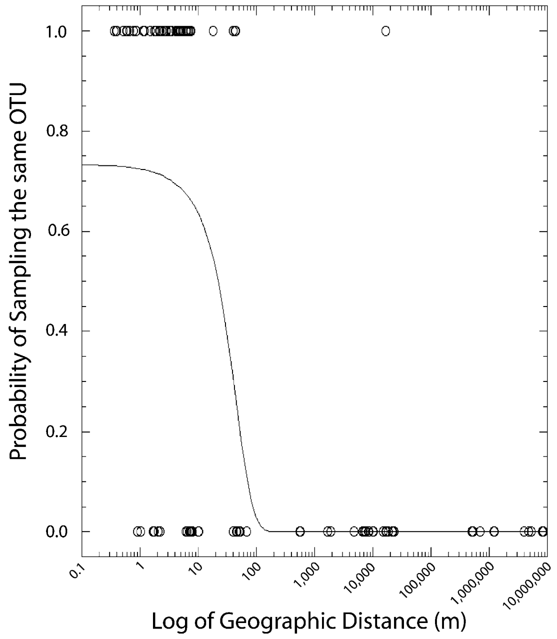

Fig. S3. The probability curve of sampling any one specific OTU (at the 97% sequence similarity cutoff) as geographic distance increases between any pair of samples. Any pair-wise comparison of samples that shared one or more OTUs was cataloged as “one,” and any pair-wise comparison of samples that shared no OTUs was cataloged as “zero.” The probability of sampling the same OTU drops to zero for almost all pair-wise comparisons beyond a geographic distance of about 100 m.

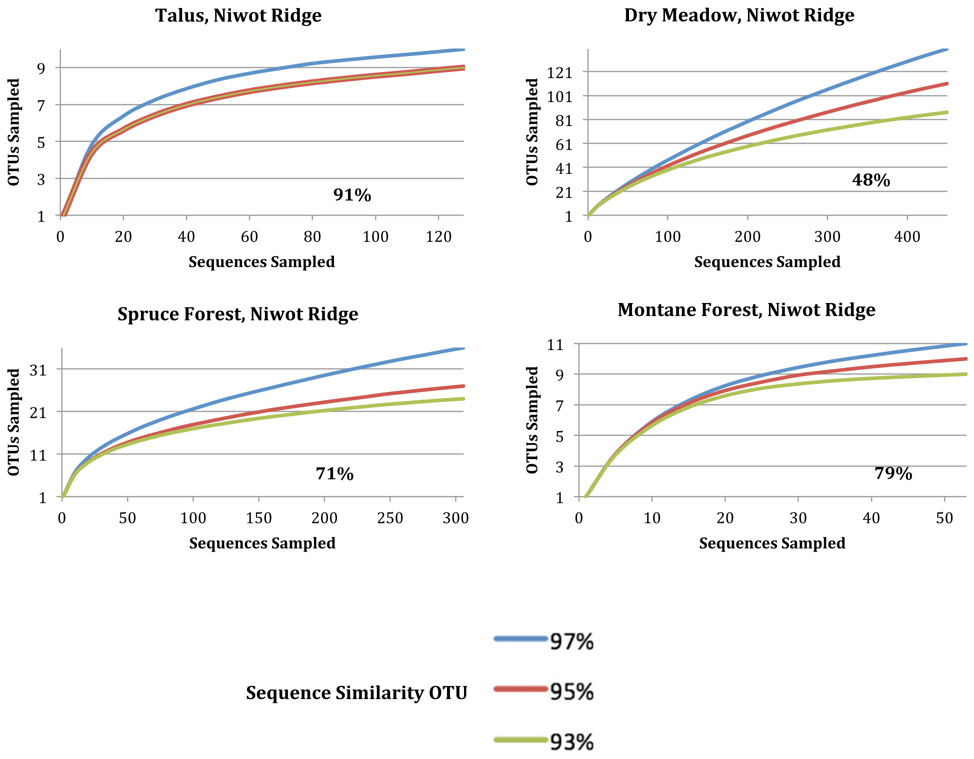

Fig. S4. Rarefaction curves for four representative sample sites including the site that was least fully sampled (dry meadow). Three OTU sequence similarity cutoffs of 97, 95, and 93% are plotted. The percent of the estimated total OTUs (at the 97% similarity level) sampled at each site (1) is included at the bottom right in each graph. All other sites had similar curves and showed that we sampled between 64 and 100% of possible OTUs at those sites. These results indicate that we did not limit sampling to only the dominant bdelloid clones of a given community.

1. Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scand J Stat* 11:265–270.

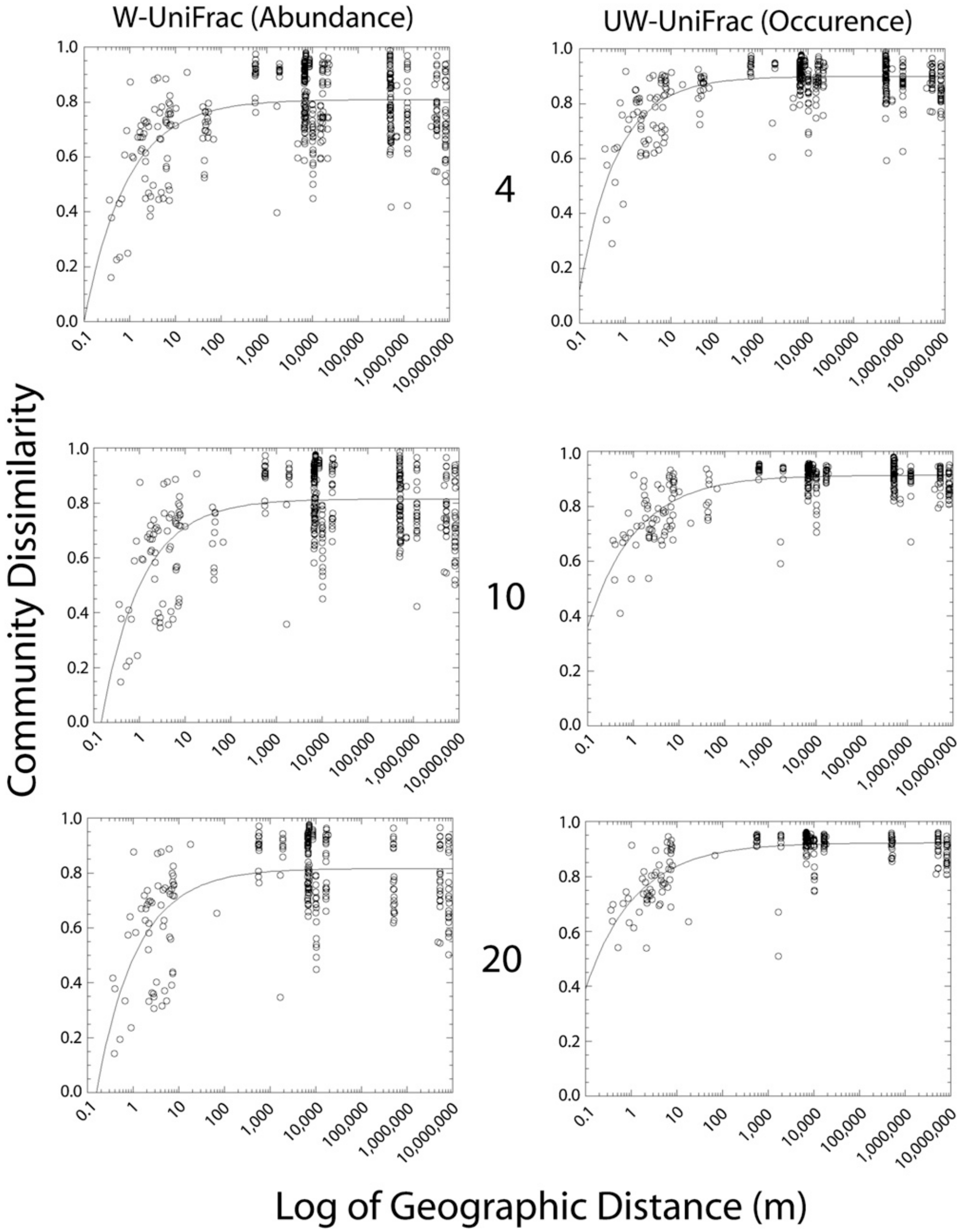

Fig. S5. Jackknife analyses of phylogenetic data using both unweighted (UW) and weighted (W) UniFrac as implemented in QIIME (1). Differing jackknife subsampling depths are shown; that is, 4, 10, or 20 sequences were subsampled (without replacement) from each geographic location. Each level of subsampling was performed for 1,000 iterations, and the means of each and all pair-wise comparisons are plotted. Local spatial autocorrelation was captured at all levels of jackknifing, confirming that at each geographic location we sampled at sufficient depth to characterize all communities.

1. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

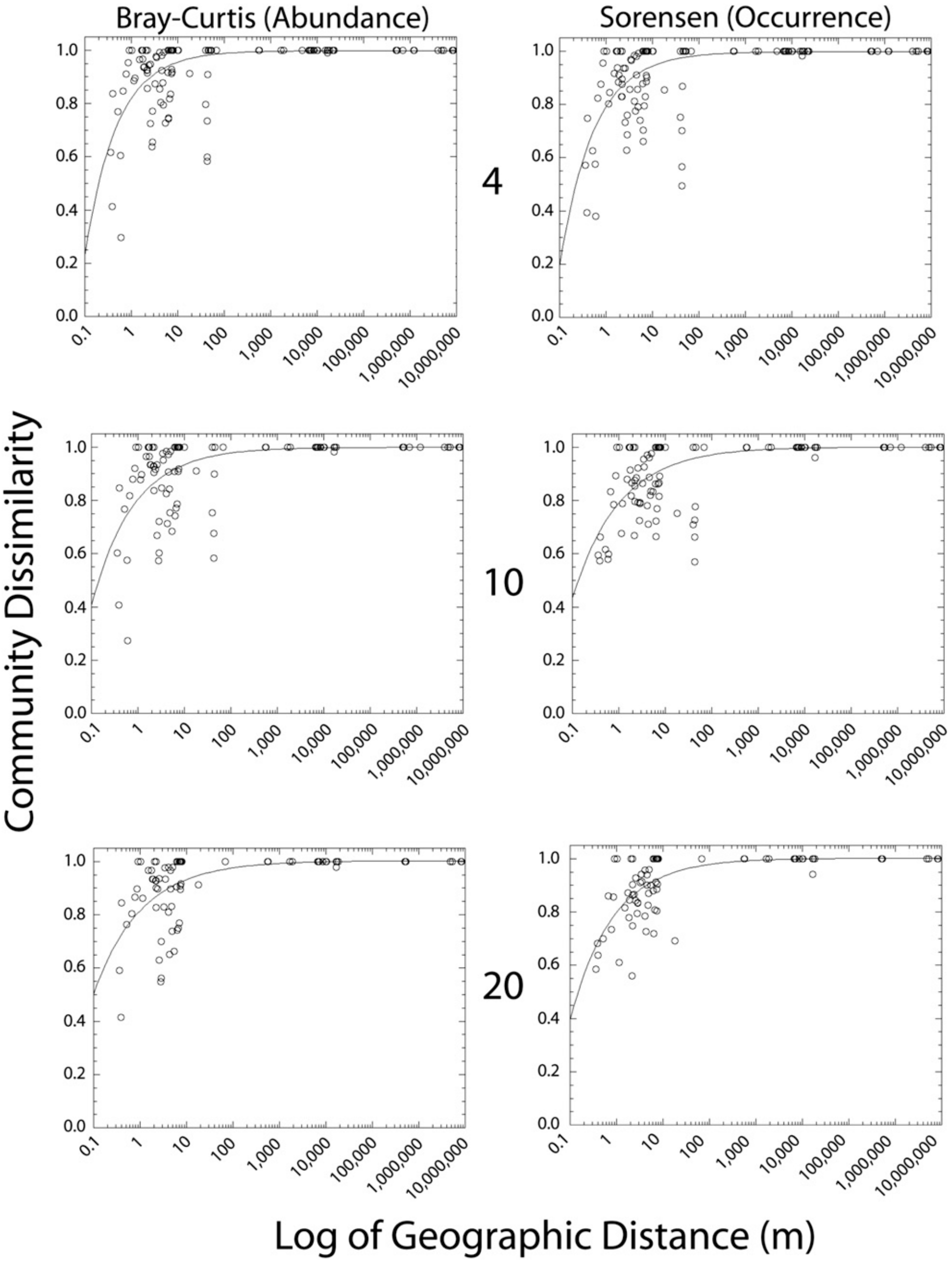

Fig. S6. Jackknife analysis of OTU-based data using both Bray–Curtis and Sorensen metrics as implemented in QIIME (1). Differing jackknife subsampling depths are shown; that is 4, 10, or 20 sequences were subsampled (without replacement) from each geographic location. Each level of subsampling was performed for 1,000 iterations, and the means of each and all pair-wise comparisons are plotted. Local spatial autocorrelation was captured at all levels of jackknifing, confirming that at each geographic location we sampled at sufficient depth to characterize all communities.

1. Caporaso JG, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336.

Table S1. GenBank accession numbers from both existing sequences and sequences generated in the present study

Table S1

C.W.B. provided nonaccessioned sequence data for use as outgroups (labeled as “not_in_gb” for “not in GenBank”). GenBank accession numbers for sequences generated in the present study are HQ174968–HQ175991.

Table S2. Mean Phylogenetic Diversity (MPD) and Net Relatedness Index (NRI) for every geographic location in the present study

Table S2

Higher values of NRI (low MPD) indicate phylogenetically constrained communities, and low NRI (high MPD) indicates more even communities (1, 2). Community designations and geospatial coordinates also are listed along with the number of sequences obtained for each geographic location.

1. Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenetics and community ecology. *Annu Rev Ecol Syst* 33:475–505.
2. Webb CO (2000) Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *Am Nat* 156:145–155.

Dataset S1. Reference phylogeny showing the placement of environmental sequences obtained in this study within the Bdelloidea

Dataset S1

This Newick-formatted cladogram can be copied and opened in tree-viewing programs such as Dendroscope (1). The tree is a majority-rule consensus tree of 1,000 maximum-likelihood bootstrap searches produced from RAxML (2) using the GTR + I + G model of evolution as chosen by MultiPhyl Online (3) and partitioned by codon position. Bdelloid and outgroup cytochrome c oxidase subunit 1 (*cox1*) sequences from are from GenBank. (Table S1 provides accession numbers.)

1. Huson DH, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460.
2. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22:2688–2690.
3. Keane TM, Naughton TJ, McInerney JO (2007) MultiPhyl: A high-throughput phylogenomics webserver using distributed computing. *Nucleic Acids Res* 35(Web Server issue):W33–37.

Dataset S2. Midpoint-rooted bdelloid-only dendrogram used for UniFrac analyses

Dataset S2

This midpoint rooted Newick-formatted tree file can be copied and opened in tree-viewing programs such as Dendroscope (1). The best-scoring maximum-likelihood inference was obtained after performing 100 full-alignment inferences from RAxML (2) using the GTR + I + G model of evolution as chosen by MultiPhyl Online (3) and partitioned by codon position. This tree was used to perform UniFrac analyses. Only sequence data produced from this study are present in the tree. (Tables S1 and S2 provide accession numbers and sample designations, respectively.)

1. Huson DH, et al. (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460.
2. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22:2688–2690.
3. Keane TM, Naughton TJ, McInerney JO (2007) MultiPhyl: A high-throughput phylogenomics webserver using distributed computing. *Nucleic Acids Res* 35(Web Server issue):W33–37.