Supplementary Figures

Title: Assessing and improving methods used in OTU-based approaches for 16S rRNA gene sequence analysis

Authors: Patrick D. Schloss & Sarah L. Westcott

Figure S1. The cumulative fraction of taxa that had a specified maximum intra-taxon distance (A) and total branch length (B) for each taxonomic level when analyzing V13 sequences. At each taxonomic level, sequences that did not affiliate with a known lineage (i.e. incertae sedis) were excluded. The numbers in parentheses next to the name of each taxonomic level indicate the number of taxa within that level that we observed.

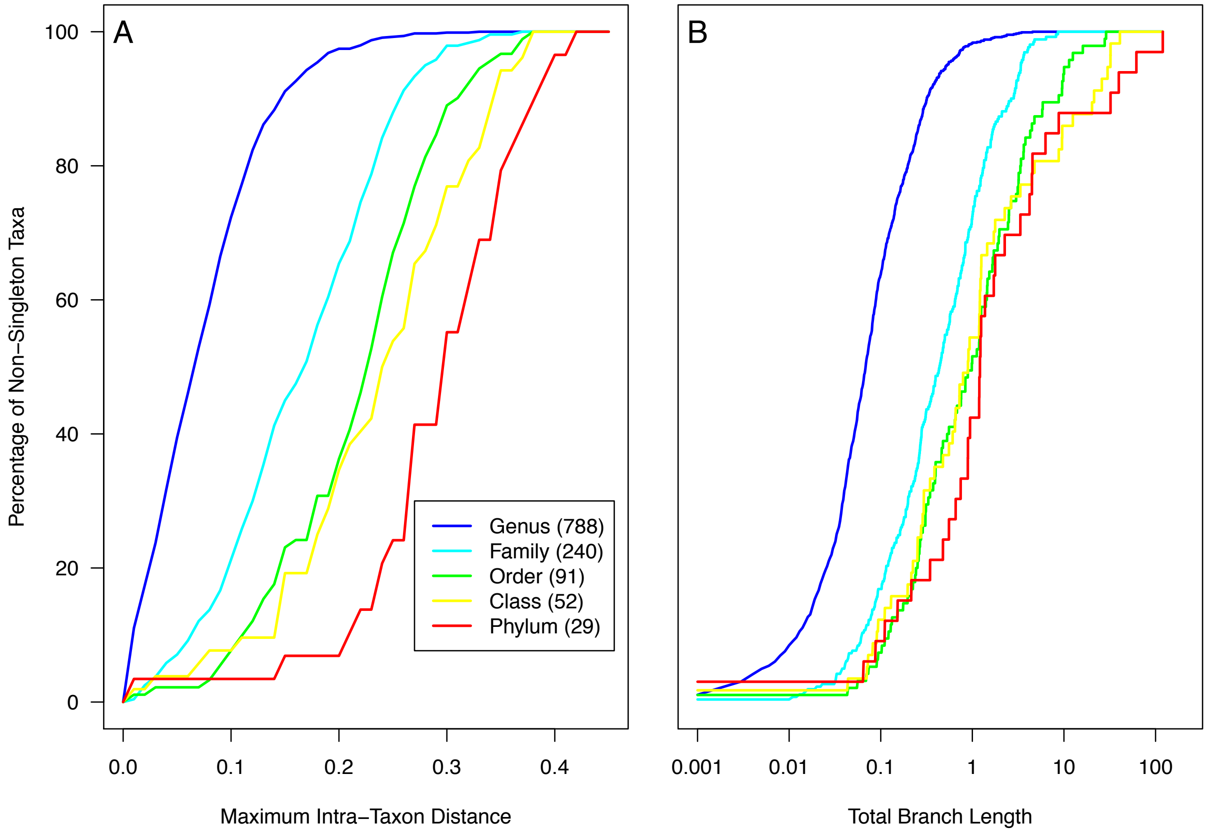

Figure S2. The cumulative fraction of taxa that had a specified maximum intra-taxon distance (A) and total branch length (B) for each taxonomic level when analyzing V35 sequences. At each taxonomic level, sequences that did not affiliate with a known lineage (i.e. incertae sedis) were excluded. The numbers in parentheses next to the name of each taxonomic level indicate the number of taxa within that level that we observed.

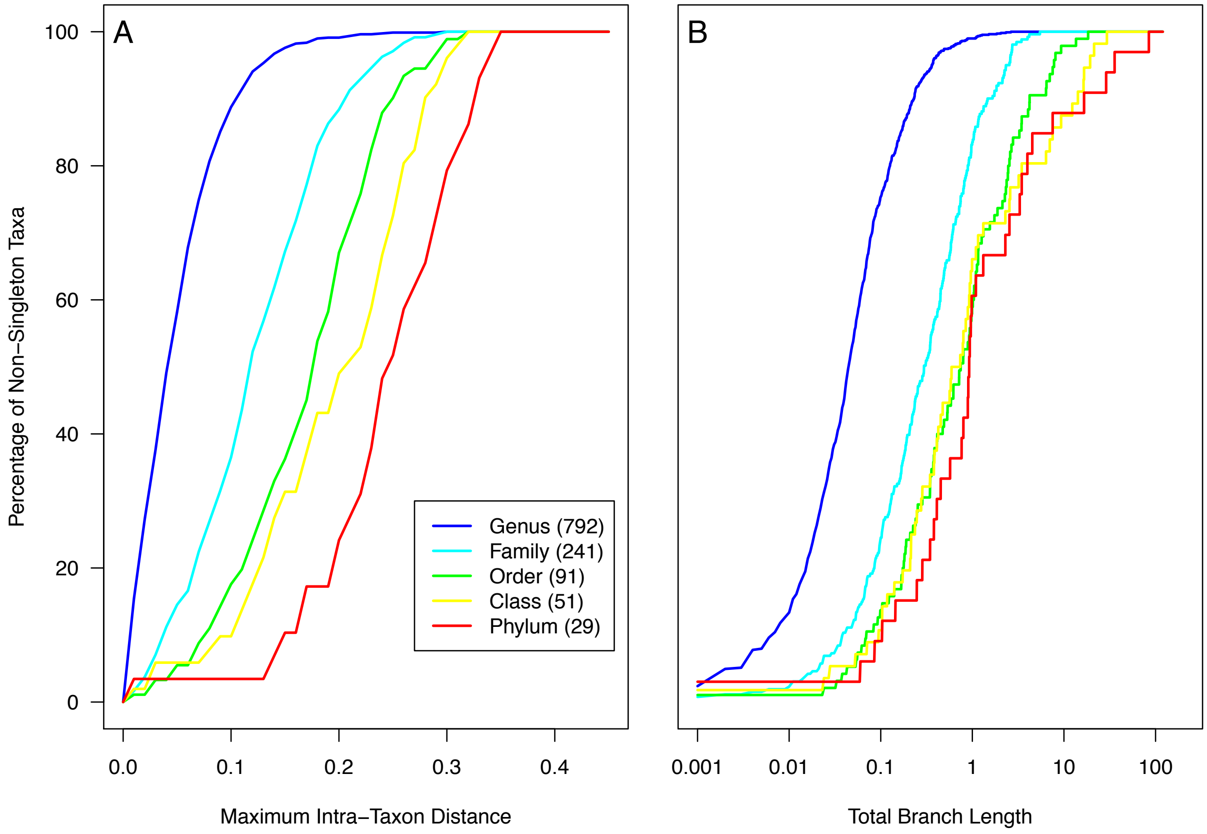

Figure S3. The variation in the Matthew's correlation coefficient calculated for OTUs identified using eight classification algorithms at genetic distances varying between 0.00 and 0.10 for V13 sequences.

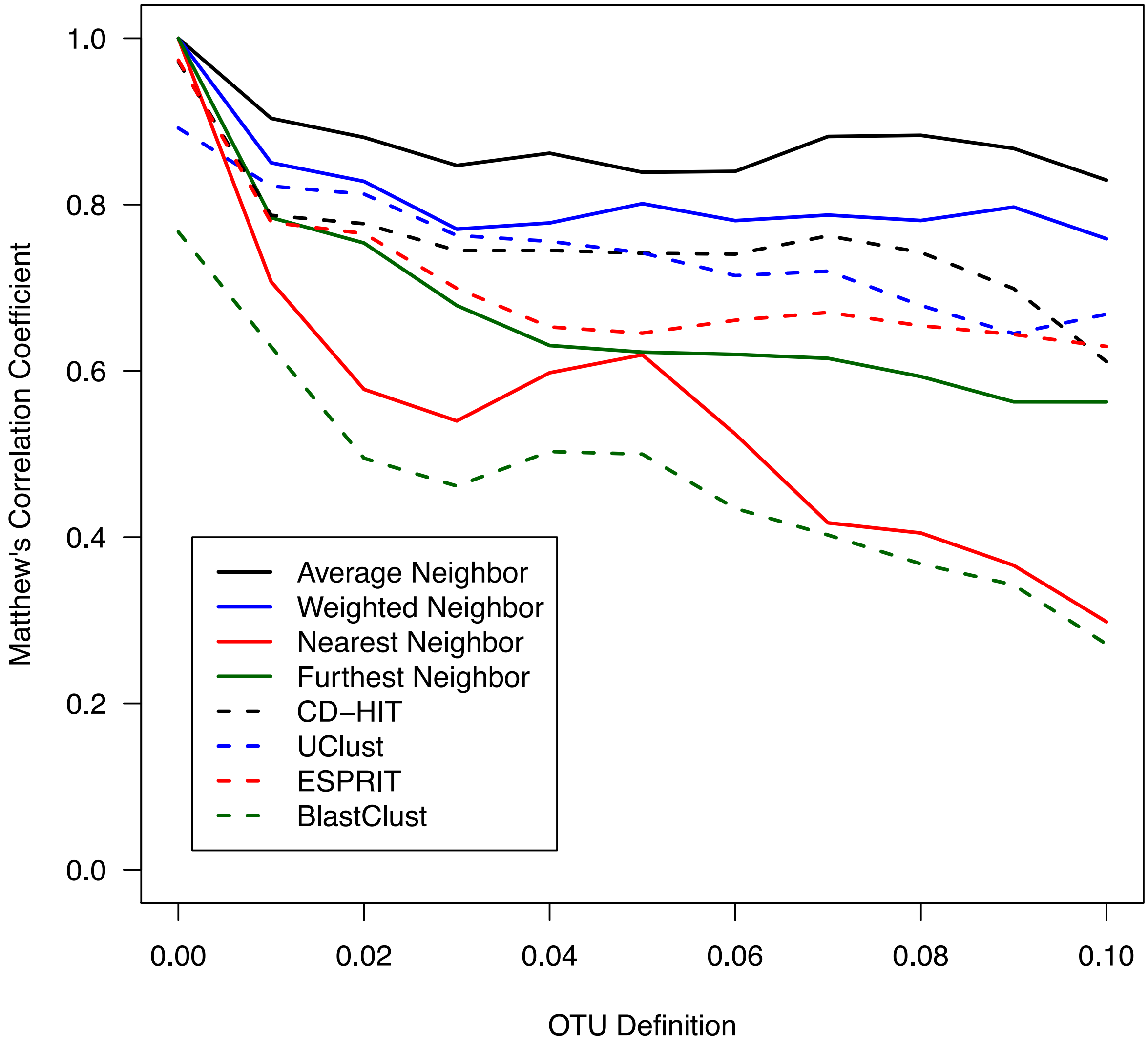

Figure S4. The variation in the Matthew's correlation coefficient calculated for OTUs identified using eight classification algorithms at genetic distances varying between 0.00 and 0.10 for V35 sequences.

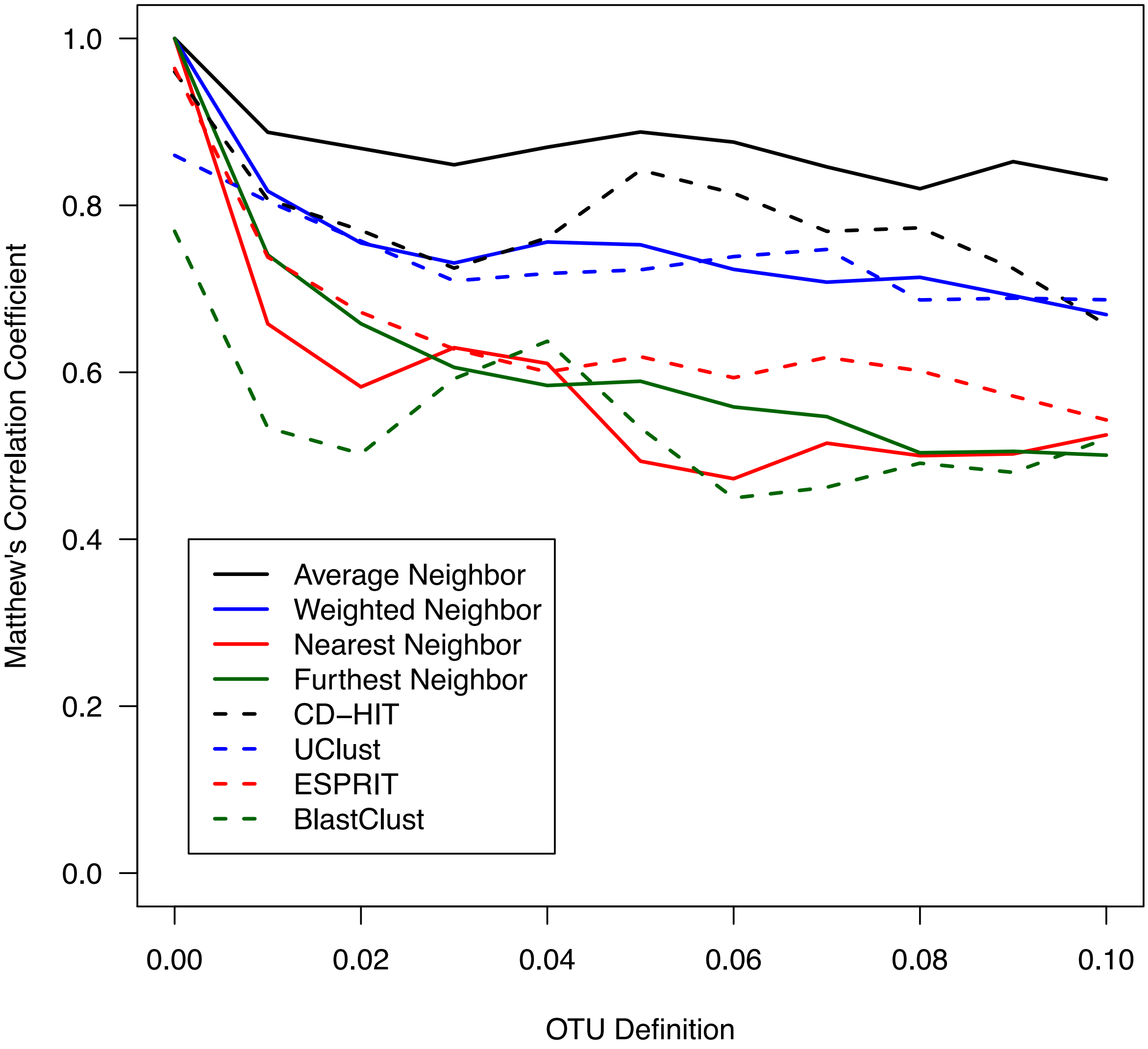

Figure S5. Comparison of the Matthew's correlation coefficients for OTUs calculated from a threshold of 0.00 to 0.10 when using the phylotype-OTU heuristic for V13 sequences. For each region, cutoff, and taxonomic level used to split the sequences, the correlation coefficients overlapped with each other except for the family and genus taxonomic levels.

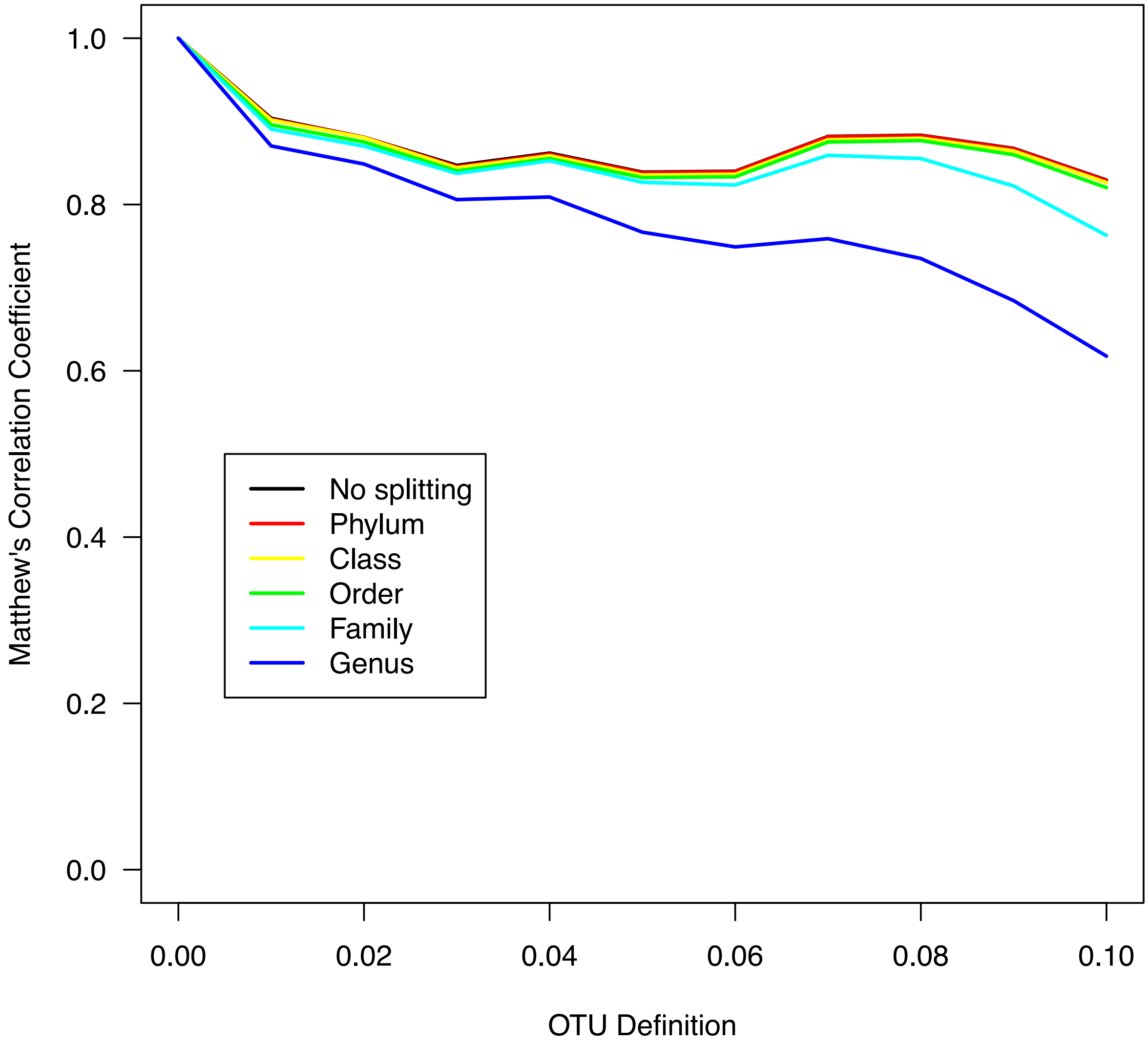

Figure S6. Comparison of the Matthew's correlation coefficients for OTUs calculated from a threshold of 0.00 to 0.10 when using the phylotype-OTU heuristic for V35 sequences. For each region, cutoff, and taxonomic level used to split the sequences, the correlation coefficients overlapped with each other except for the family and genus taxonomic levels.