

Appendix S1.

Searchable fasta files for all 12-mer through 17-mer matches to the chloroplast and mitochondrial PNA sequence clamps can be found here: <https://github.com/sjackrel/Identifying-the-plant-associated-microbiome-across-aquatic-and-terrestrial-environments>

Note that the pPNA14merD.fna includes all OTUs with an exact match of 14 (GGCTCAACCCTGGA) of the 17 (GGCTCAACCCTGGACAG) base pairs of the pPNA chloroplast blocking primer. From the entire GreenGenes Database containing 99,322 there were 1405 matches. This includes mostly Proteobacteria (1069 OTUs), Chloroplast (67 OTUs), other Cyanobacteria (23 OTUs), Firmicutes (78 OTUs), Actinobacteria (53 OTUs), Bacteroidetes (29 OTUs), and Verrucomicrobia (27 OTUs).

Appendix S2.

Figure 1. Procrustes analysis illustrating distance in principal coordinate space among microbial communities of each soil sample. Distance between replicate samples amplified with the EMP versus EMP-PNA method shown with connecting lines. Sample # 11 is from beneath a tree growing furthest downstream on the Hoko River ($48^{\circ} 15'29.58$ N, $124^{\circ} 21'8.59$ W). Sample numbers increase for trees growing further upstream. White lines point to the EMP sample and red lines point to the corresponding PNA sample.

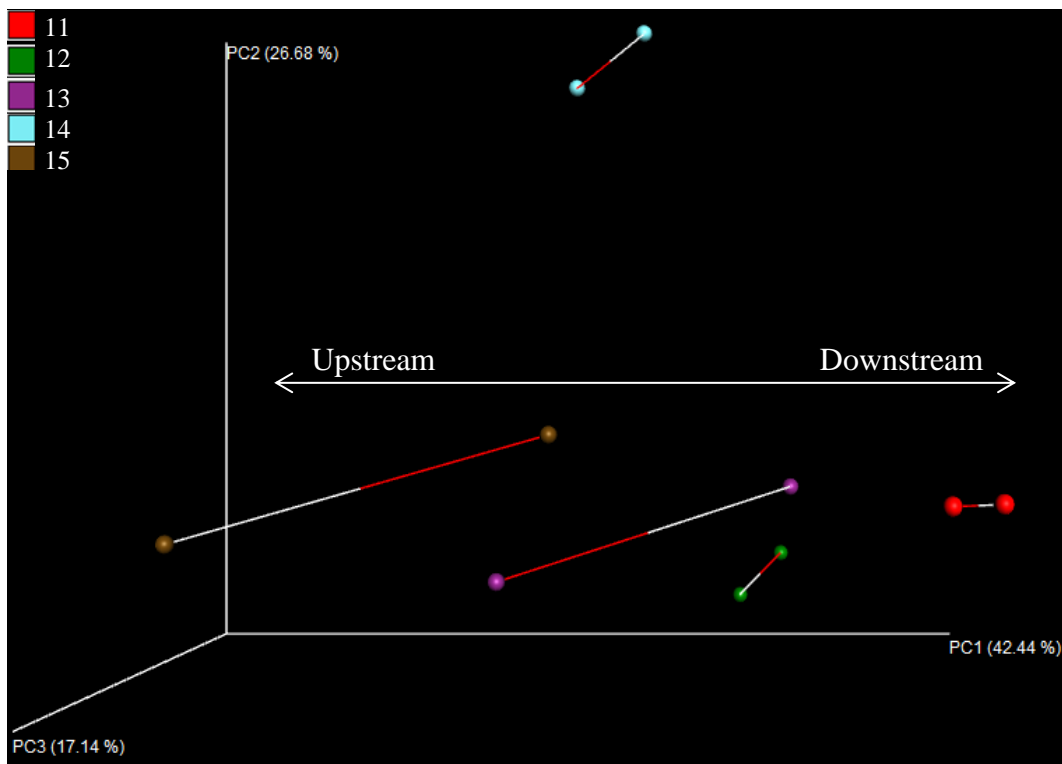


Figure 2. Procrustes analysis illustrating distance in PC space among microbial communities of each freshwater sample. Distance between replicate samples amplified with the EMP versus EMP-PNA method shown with connecting lines. Samples were collected upstream of two deployment sites on the Hoko River and two deployment sites on the Sekiu River. White lines point to the EMP sample and red lines point to the corresponding PNA sample.

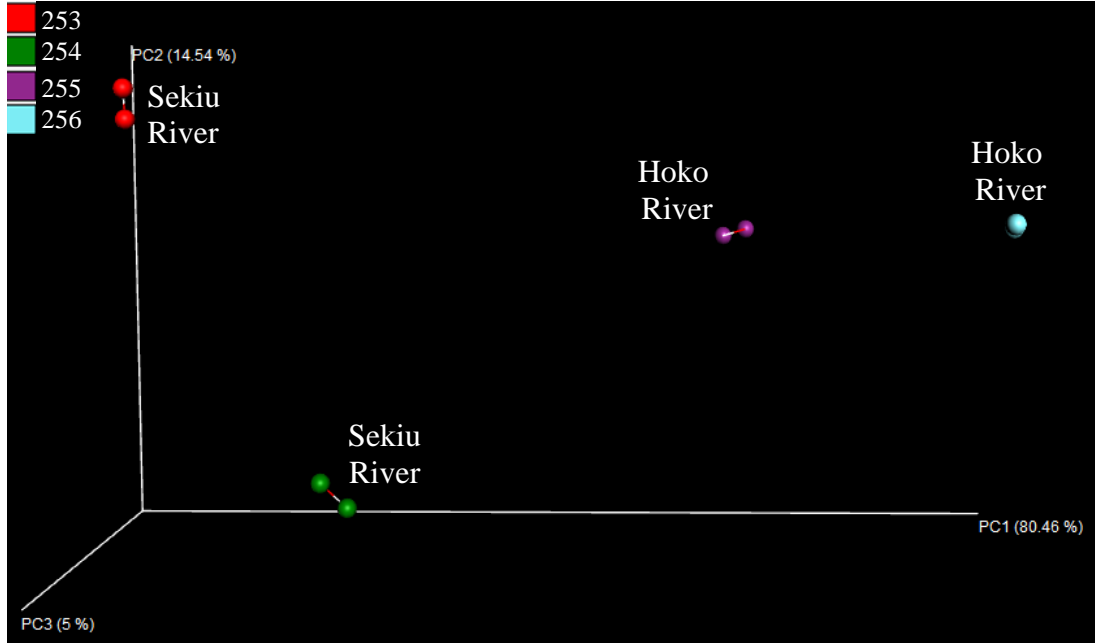


Figure 3. Procrustes analysis illustrating distance in PC space among microbial communities of each seawater sample. Distance between replicate samples amplified with the EMP versus EMP-PNA method shown with connecting lines. Sample #1501 – 1509 and #1514 – 1519 are surface samples at Slip Point, WA (48.26° N, 124.25° W); #1520 – 1523 are surface samples at Tatoosh Island, WA (48.39° N, 124.74° W), #1511 is at 100 m deep, #1524 is at 125 m deep, #1510 is at 140 m deep, #1512 is at 325 m deep, and #1513 is at 300 m deep. White lines point to the EMP sample and red lines point to the corresponding PNA sample.

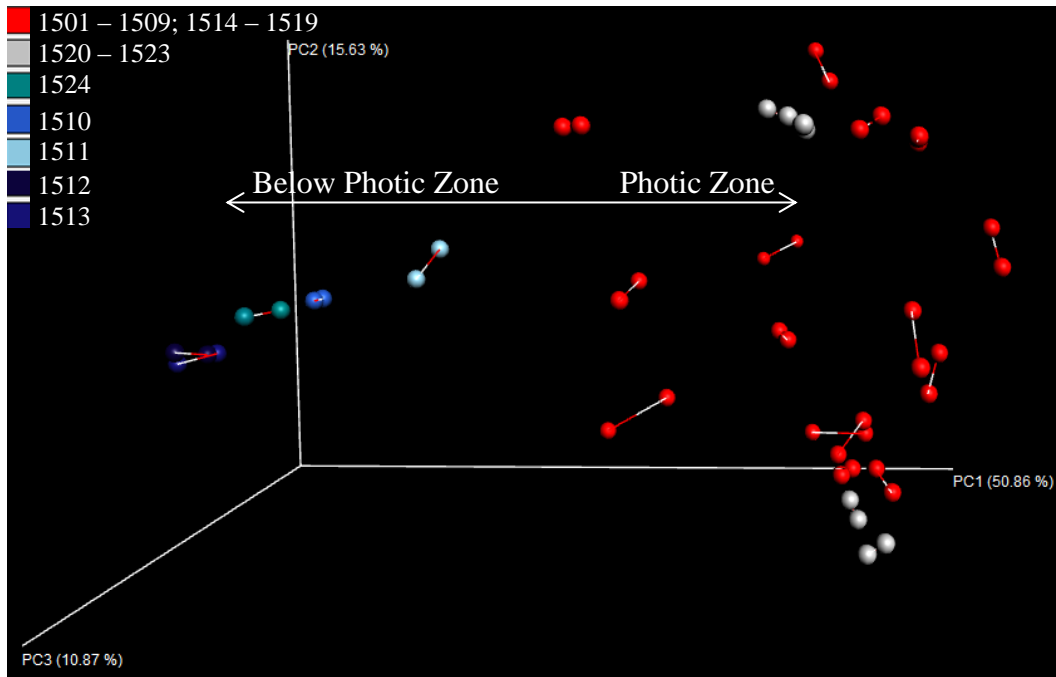


Figure 4. Procrustes analysis illustrating distance in PC space among microbial communities of each aquatic leaf sample. Distance between replicate samples amplified with the EMP versus EMP-PNA method shown with connecting lines. Sample #55 – 58 are leaves taken from the same red alder tree alongside the Sekiu River, but deployed in different locations. Sample #107 – 110 are leaves take from the same red alder tree growing alongside the Hoko River, but deployed in different locations. White lines point to the EMP sample and red lines point to the corresponding PNA sample.

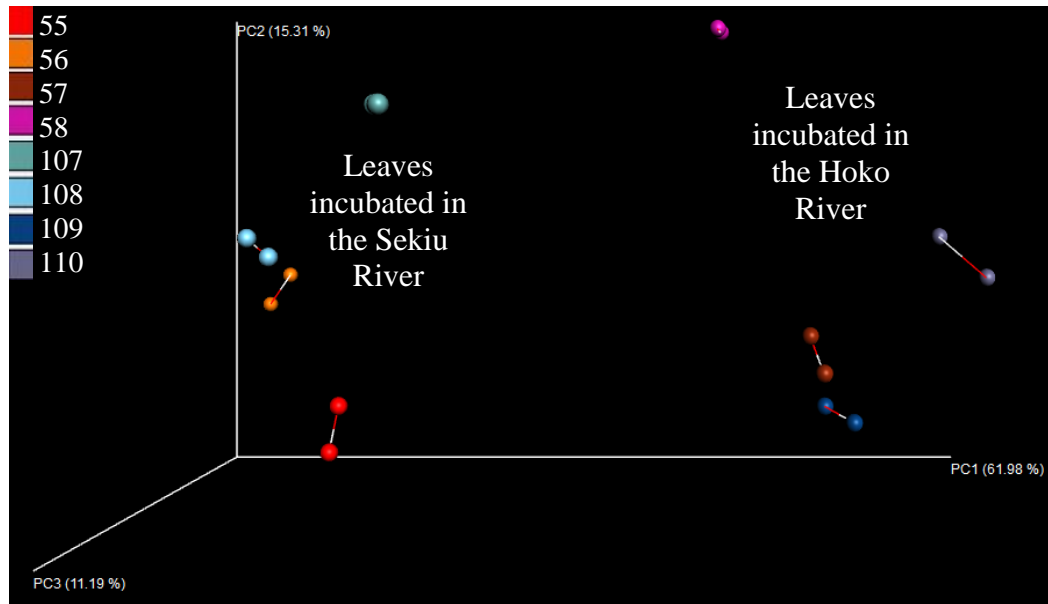
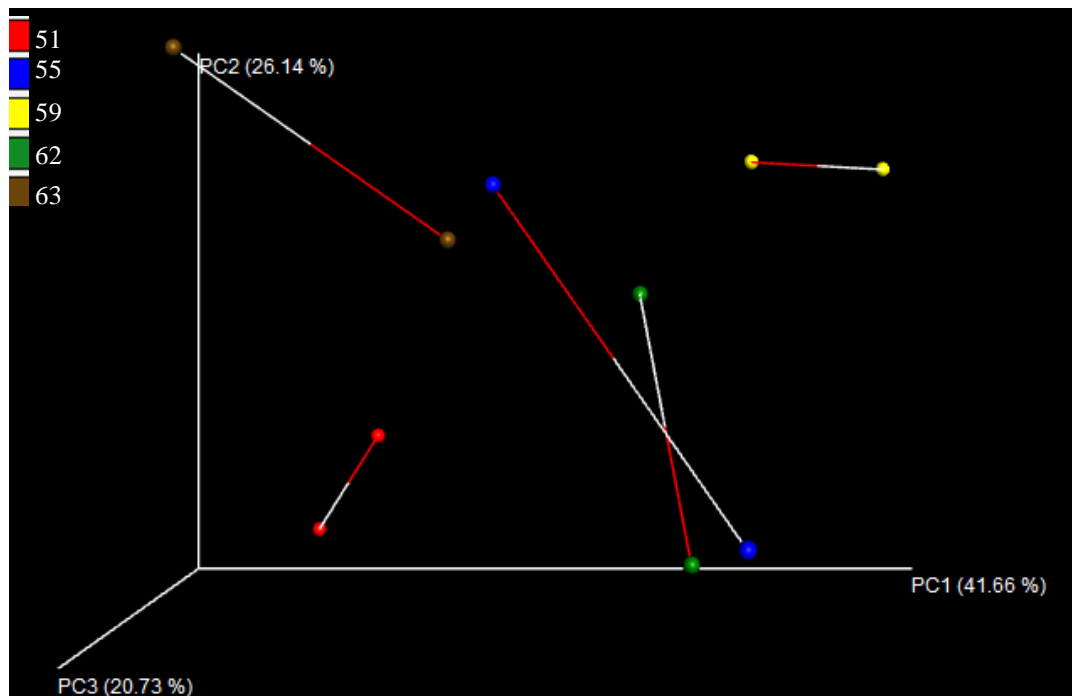


Figure 5. Procrustes analysis illustrating distance in PC space among microbial communities of each terrestrial leaf sample. Distance between replicate samples amplified with the EMP versus EMP-PNA method shown with connecting lines. Each sample consisted of one leaf from an individual red alder tree. Increasing sample number indicates that the parent tree was growing further upstream alongside the Hoko River. White lines point to the EMP sample and red lines point to the corresponding PNA sample.



Appendix S3.

Figure 1. OTU #, Phylogenetic, and Chao's alpha diversity metrics (mean + SE) of all samples by environment type (terrestrial leaves, aquatic leaves, freshwater, seawater, or soil samples) amplified with the EMP vs EMP-PNA method. (B) indicates filtering out only chloroplast and mitochondria. (A) indicates filtering out chloroplast, mitochondria, and OTUs in Appendix S1.

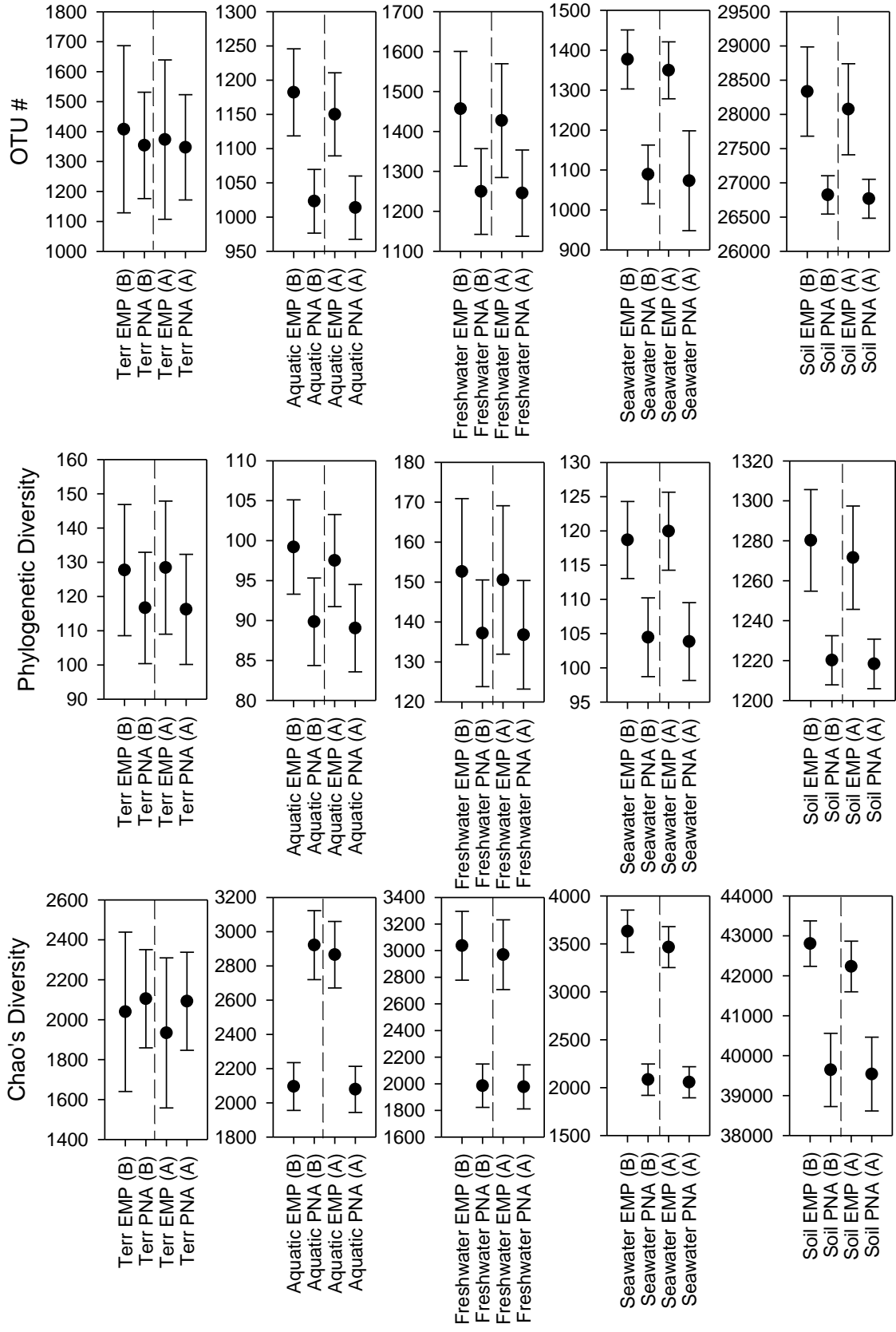


Figure 2. Rarefaction curves illustrating alpha diversity of microbial taxa via number of OTUs in aquatic leaves (A), freshwater (B), seawater (C), soil (D), and terrestrial leaf (E) samples sequenced either via the chloroplast and mitochondria-blocking EMP-PNA method (Blue) or EMP method (Red). Filtering out only chloroplast and mitochondria (not OTUs in Appendix S1).

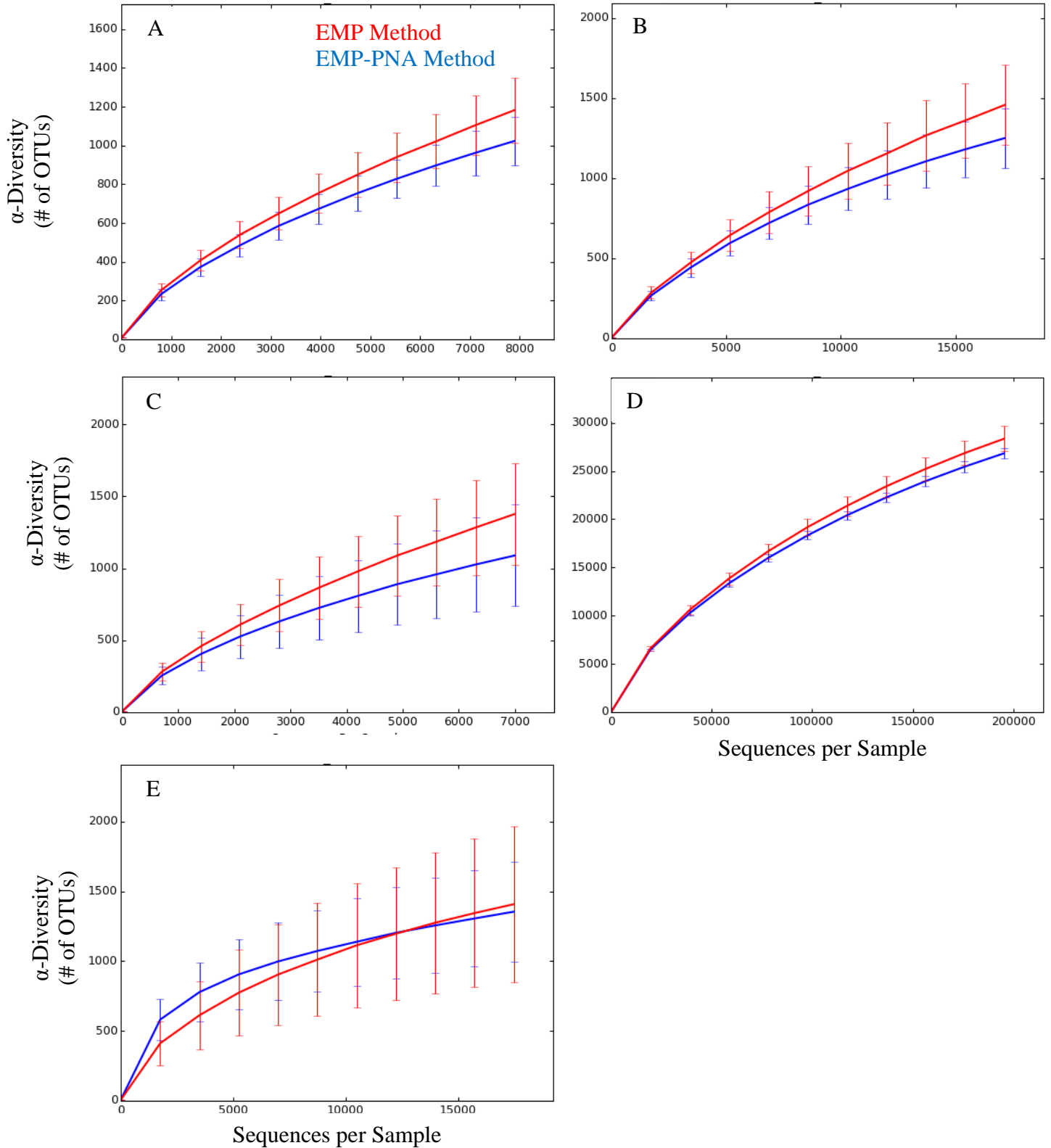
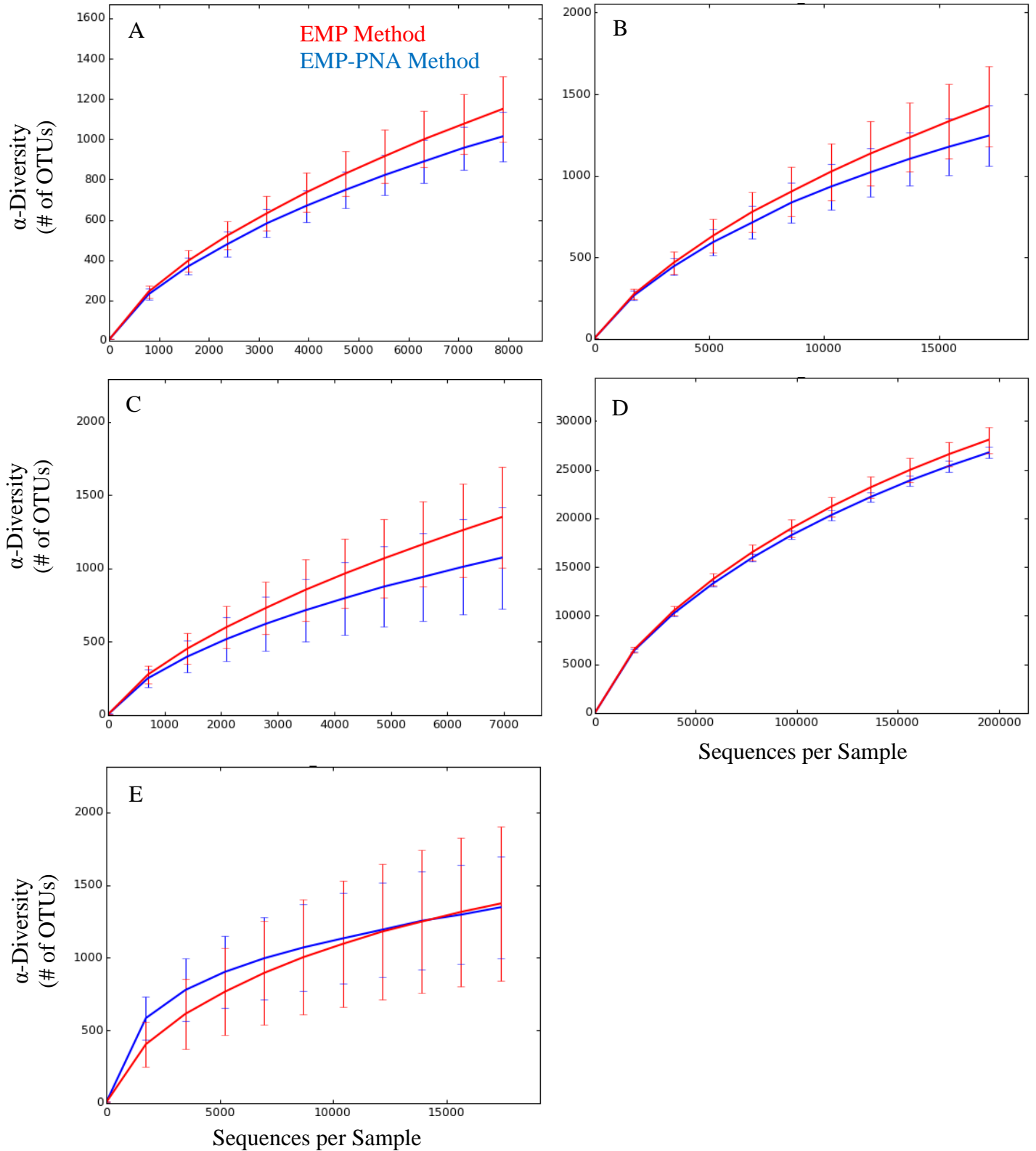


Figure 3. Rarefaction curves illustrating alpha diversity of microbial taxa via number of OTUs in aquatic leaves (A), freshwater (B), seawater (C), soil (D), and terrestrial leaf (E) samples sequenced either via the chloroplast and mitochondria-blocking EMP-PNA method (Blue) or EMP method (Red). Filtering out chloroplast, mitochondria, and OTUs in Appendix S1.



Appendix S4.

Figure 1. Paired comparisons of freshwater samples sequenced via the EMP method versus the organelle-blocking EMP-PNA method. Relative abundance of microbial taxa at the family level depicted via color. 'Before' samples depict communities after filtering out chloroplast and mitochondrial sequences. 'After' samples depict communities after additionally filtering out OTUs in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity as a measure of discontinuity by amplification method.

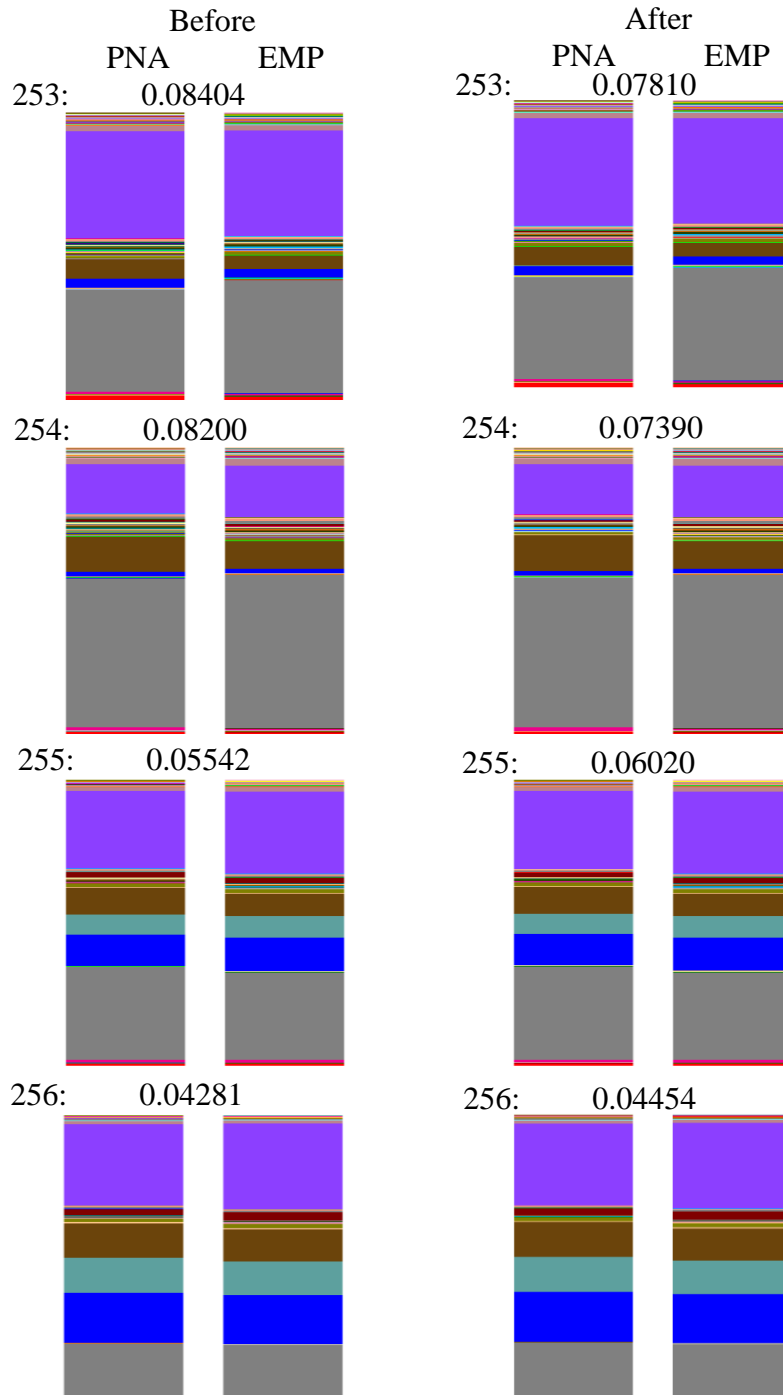


Table 1. Microbial taxa at lower relative abundance in freshwater samples when sequenced via the EMP-PNA method versus EMP method. The first column lists rank order abundance of each taxa in the entire freshwater sample set. Reported p-values are from paired t-tests with and without false discovery rate correction. Values in () are p-values from Wilcoxon sign-rank tests.

Abundance	P-value	FDR	Taxonomic Classification
# 306	0.0039 (0.125)	1 (1)	Proteobacteria;c_Betaproteobacteria;o_Methylophilales;f_;g_
# 7	0.014 (0.125)	1 (1)	Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Microbacteriaceae;g_Candidatus Rhodoluna
# 62	0.016 (0.125)	1 (1)	OD1;c_ZB2;o_;f_;g_
# 4	0.020 (0.125)	1 (1)	Bacteroidetes;c_Flavobacteriia;o_Flavobacteriales;f_Flavobacteriaceae;g_Flavobacterium
# 328	0.023 (0.125)	1 (1)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Hyphomicrobiaceae;g_Pedomicrobium
# 259	0.027 (0.125)	1 (1)	Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Comamonadaceae;g_Acidovorax
# 202	0.031 (0.125)	1 (1)	Actinobacteria;c_Rubrobacteria;o_Rubrobacterales;f_Rubrobacteraceae;g_Rubrobacter
# 53	0.031 (0.125)	1 (1)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_
# 89	0.032 (0.125)	1 (1)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Phyllobacteriaceae;g_
# 30	0.036 (0.125)	1 (1)	Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_;g_
# 35	0.034 (0.125)	1 (1)	Proteobacteria;c_Deltaproteobacteria;o_Myxococcales;f_;g_
# 33	0.037 (0.125)	1 (1)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Hyphomonadaceae;g_
# 153	0.038 (0.125)	1 (1)	Parvarchaeota;c_[Parvarchaea];o_YLA114;f_;g_

Figure 2. Paired comparisons of soil samples sequenced via the EMP method versus the organelle-blocking EMP-PNA method. Relative abundance of microbial taxa at the family level depicted via color. ‘Before’ samples depict communities after filtering out chloroplast and mitochondrial sequences. ‘After’ samples depict communities after additionally filtering out OTUs in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity as a measure of discontinuity by amplification method.

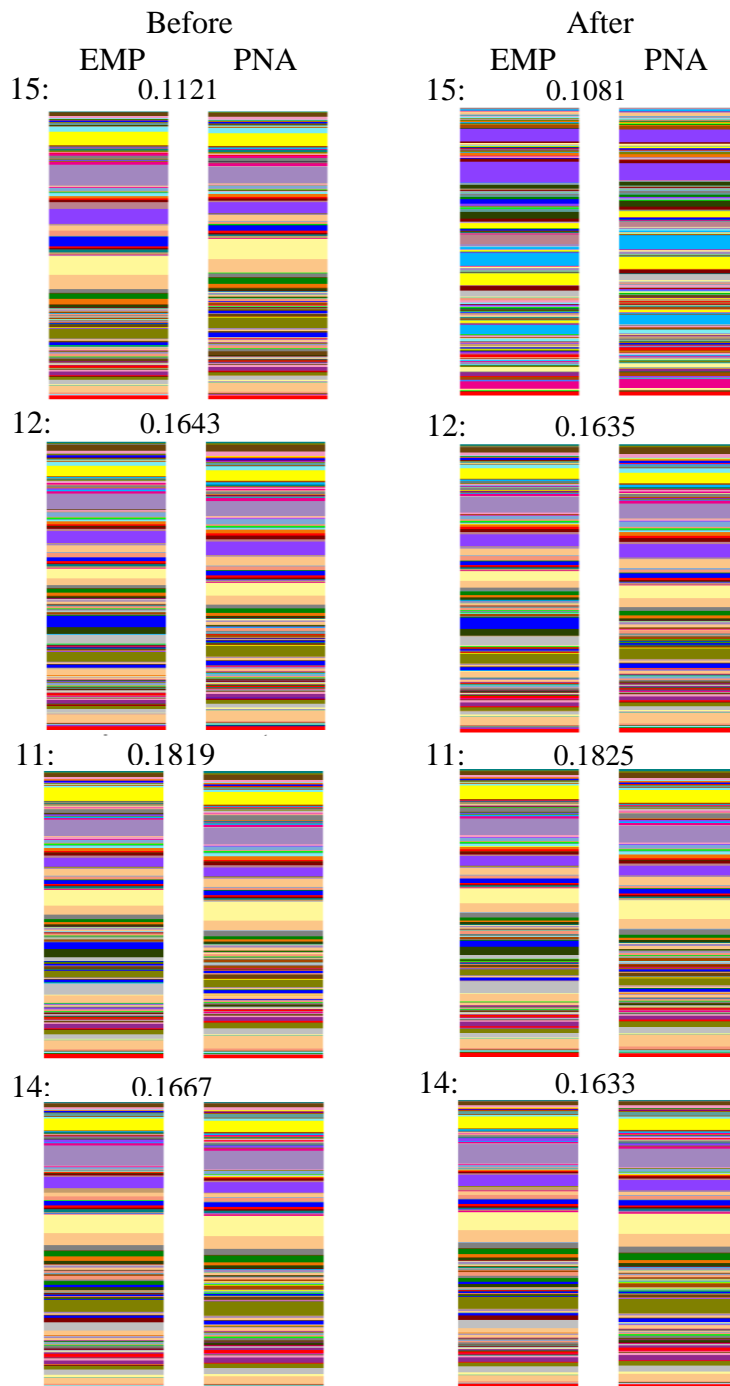


Table 2. Microbial taxa at lower relative abundance in soil samples when sequenced with the EMP-PNA method versus EMP method. The first column lists rank order abundance of each taxa in the entire soil sample set. Reported p-values are from paired t-tests with and without false discovery rate correction. Values in () are p-values from Wilcoxon sign-rank tests.

Abund.	P-value	FDR	Taxonomic Classification
# 1167	0.0011 (0.125)	0.24 (0.655)	Verrucomicrobia;c_Spartobacteria;o_Chthoniobacterales;f_Chthoniobacteraceae
# 684	0.0036 (0.125)	0.24 (0.655)	Planctomycetes;c_C6;o_d113
# 672	0.0089 (0.125)	0.24 (0.655)	OP3;c_koll11
# 785	0.010 (0.125)	0.25 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_Amaricoccus
# 866	0.011 (0.125)	0.25 (0.655)	Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Comamonadaceae;g_Rhodoferax
# 624	0.011 (0.125)	0.50 (0.655)	GN02;c_BB34;o_;f_;g_
# 730	0.013 (0.125)	0.24 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Caulobacterales;f_Caulobacteraceae;Other
# 824	0.014 (0.125)	0.28 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Sphingomonadales;f_Sphingomonadaceae;g_Novosphingobium
# 783	0.015 (0.125)	0.35 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Hyphomonadaceae;g_
# 100	0.016 (0.125)	0.24 (0.655)	Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_ACK-M1;g_
# 768	0.017 (0.125)	0.25 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Phyllobacteriaceae;Other
# 978	0.019 (0.125)	0.25 (0.655)	Proteobacteria;c_Deltaproteobacteria;o_Myxococcales;f_OM27;g_
# 792	0.019 (0.125)	0.25 (0.655)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;Other

Figure 3. Paired comparisons of aquatic leaf samples sequenced via the EMP method versus the EMP-PNA method. Relative abundance of microbial taxa at the family level depicted via color. ‘Before’ samples depict communities after filtering out chloroplast and mitochondrial sequences. ‘After’ samples depict communities after additionally filtering out OTUs in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity as a measure of discontinuity by amplification method.

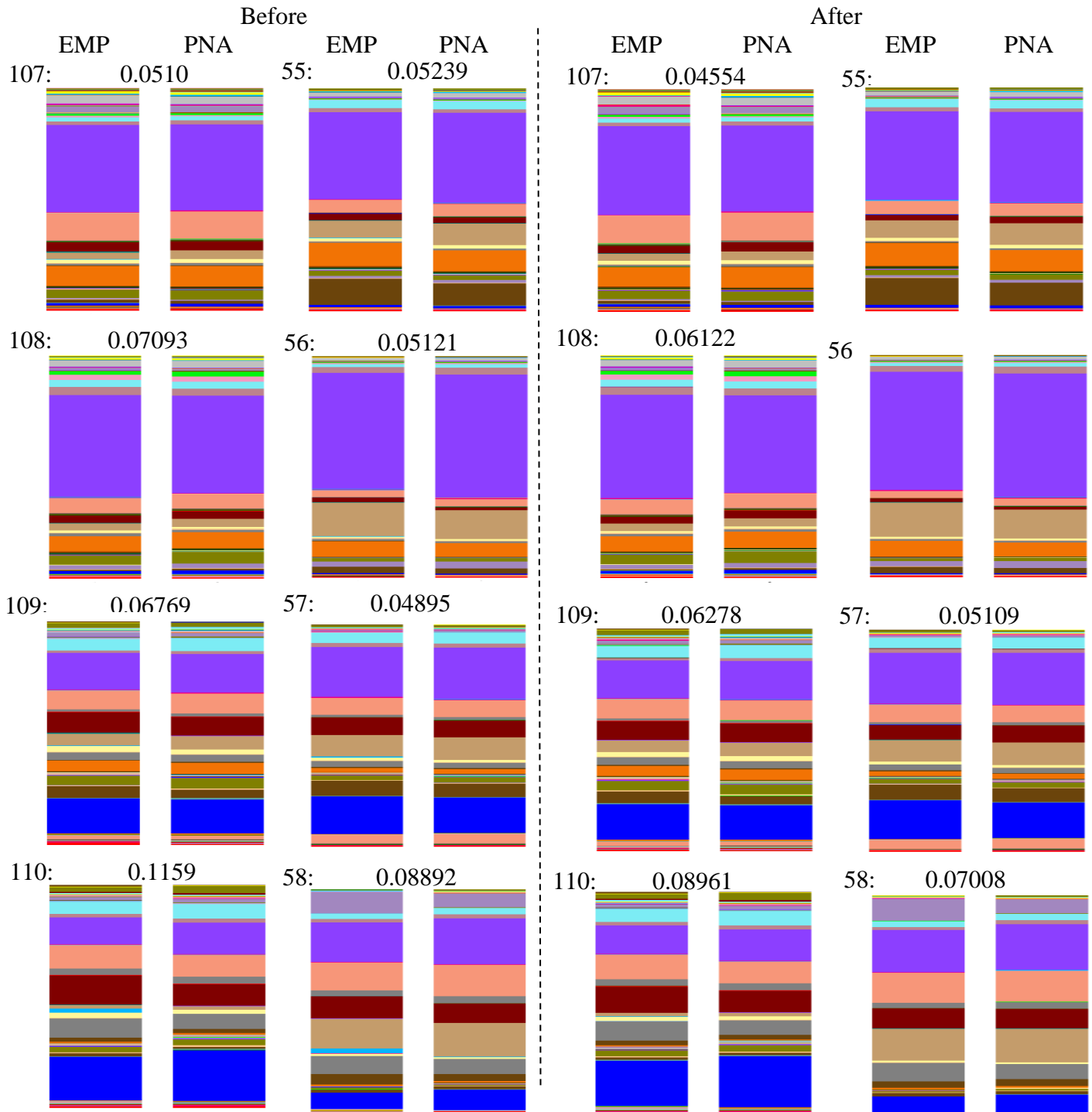


Table 3. Microbial taxa at lower relative abundance in aquatic leaf pack samples when sequenced via the EMP-PNA method versus the EMP method. The first column lists rank order abundance of each taxa in the entire aquatic leaf pack sample set. Reported p-values are from paired t-tests with and without false discovery rate correction. Values in () are p-values from Wilcoxon sign-rank tests.

Abund.	P-value	FDR	Taxonomic Classification
# 221	0.00029 (0.0078)	0.048 (0.35)	Crenarchaeota;c_Thaumarchaeota;o_Nitrososphaerales;f_Nitrososphaeraceae;g_Candidatus Nitrososphaera
# 222	0.00052 (0.0078)	0.048 (0.35)	Actinobacteria;c_Rubrobacteria;o_Rubrobacterales;f_Rubrobacteraceae;g_Rubrobacter
# 79	0.00071 (0.0078)	0.048 (0.35)	Proteobacteria;c_Deltaproteobacteria;o_Bdellovibrionales;f_Bdellovibrionaceae;g_Bdellovibrio
# 168	0.00089 (0.0078)	0.048 (0.35)	Verrucomicrobia;c_Spartobacteria;o_Chthoniobacterales;f_Chthoniobacteraceae;g_DA101
# 27	0.00097 (0.0078)	0.048 (0.35)	Unassigned;Other;Other;Other;Other;Other
# 66	0.0011 (0.0078)	0.048 (0.35)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Hyphomonadaceae
# 28	0.0013 (0.0078)	0.048 (0.35)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae
# 109	0.0023 (0.016)	0.050 (0.53)	Proteobacteria;c_Alphaproteobacteria;o_Caulobacterales;f_Caulobacteraceae;g_Phenylobacterium
# 173	0.0060 (0.0078)	0.075 (0.35)	Proteobacteria;c_Gammaproteobacteria;o_Methylococcales;f_Crenotrichaceae;g_Crenothrix
# 20	0.0091 (0.0078)	0.134 (0.35)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Hyphomicrobiaceae;g_Devosia

Figure 4. Paired comparisons of terrestrial leaf samples sequenced with the EMP method versus the organelle-blocking EMP-PNA method. Relative abundance of microbial taxa at the family level depicted via color. ‘Before’ samples depict communities after filtering out chloroplast and mitochondrial sequences. ‘After’ samples depict communities after additionally filtering out OTUs in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity as a measure of discontinuity by amplification method. Note the high variability among sample: the most abundant bacterial families in each sample are Holophagaceae (51), Rhodobacteraceae (55), Alteromonadaceae (62), and Sphingomonadaceae (63).

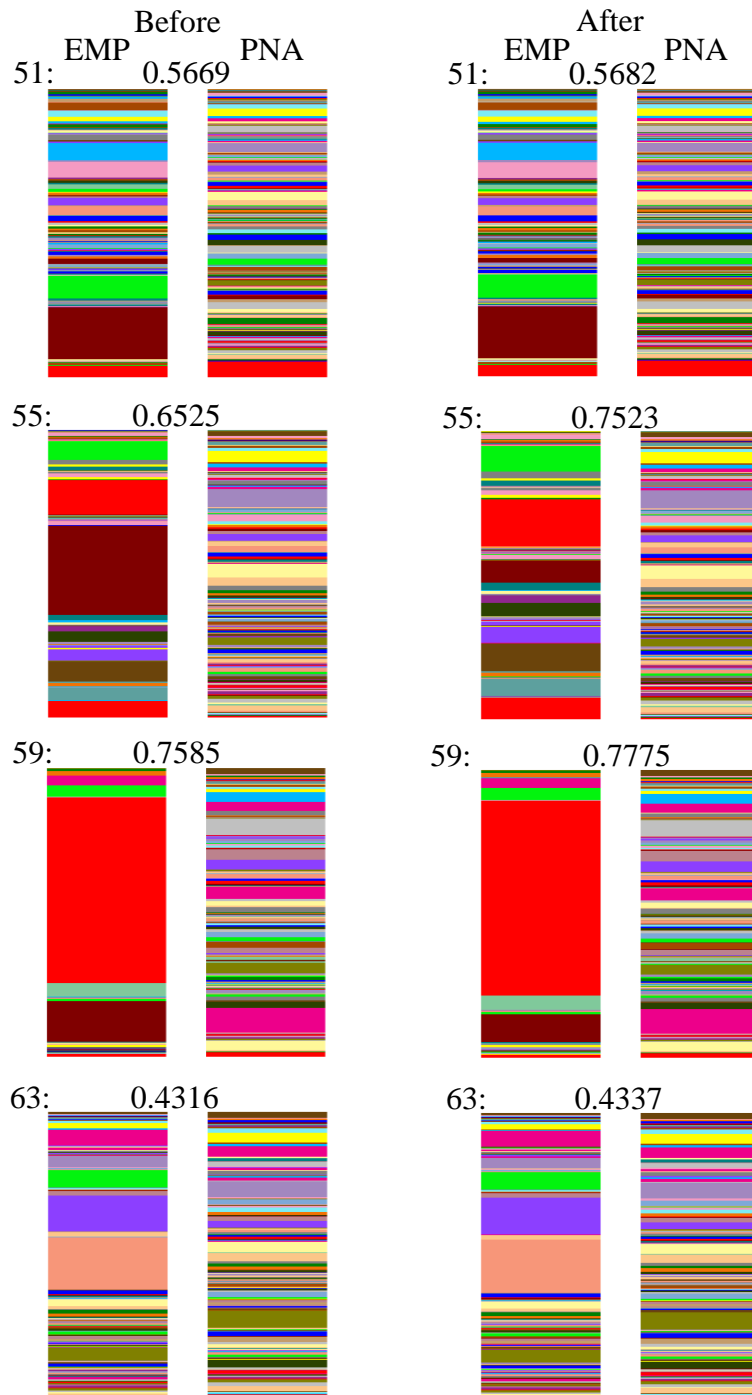


Figure 5. Seawater samples from Split Point, Washington (48.26° N, 124.25° W). Relative abundance of microbial taxa at the family level depicted via color. (A) Includes all OTUs after filtering out chloroplast and mitochondria, and (B) excludes all chloroplast, mitochondria and OTUs listed in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity.

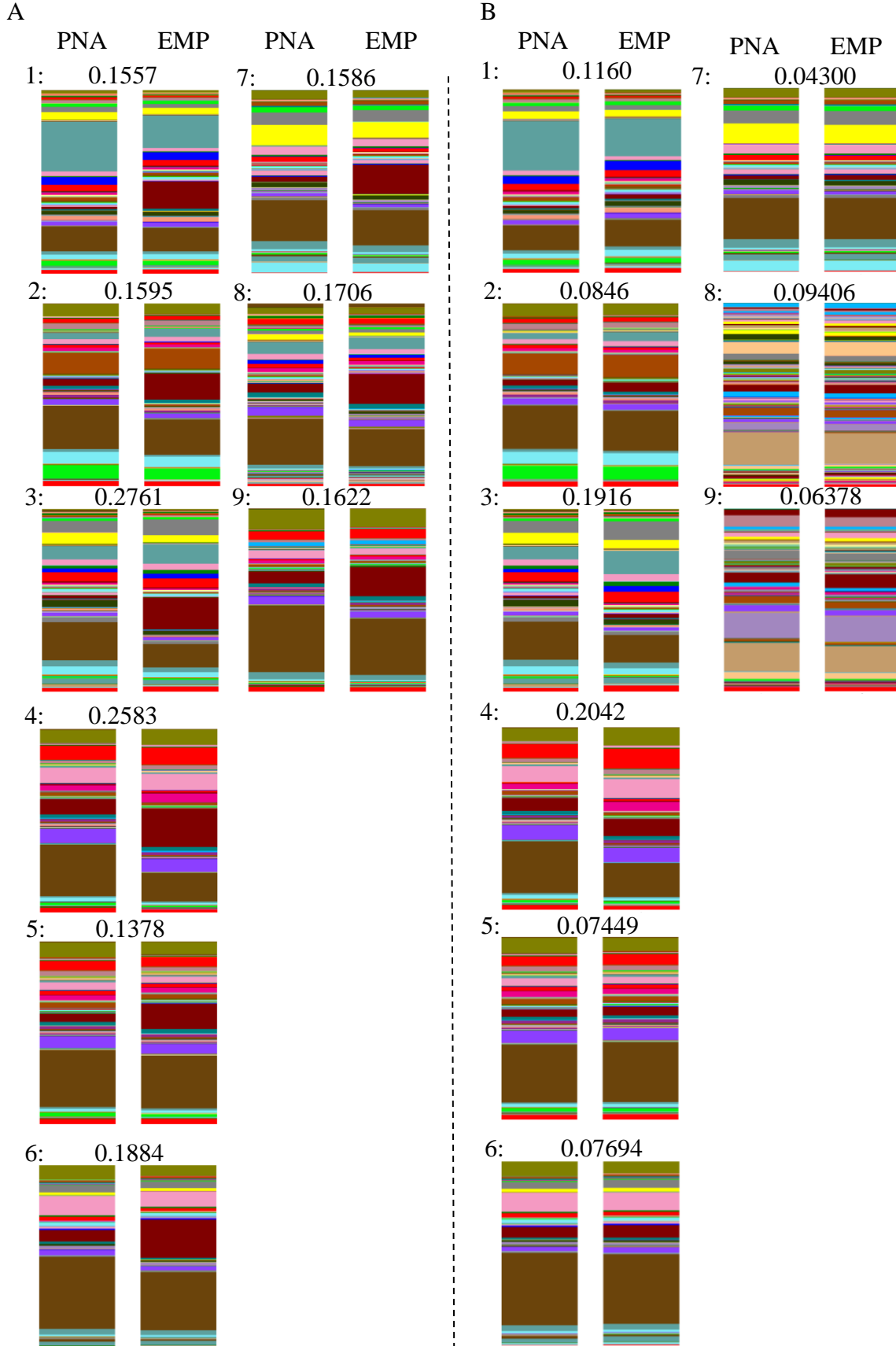


Figure 6. Seawater samples from Split Point, Washington (48.26° N, 124.25° W). Relative abundance of microbial taxa at the family level depicted via color. (A) Includes all OTUs after filtering out chloroplast and mitochondria, and (B) excludes all chloroplast, mitochondria and OTUs listed in Appendix S1. Weighted UniFrac distances between replication samples quantify community similarity.

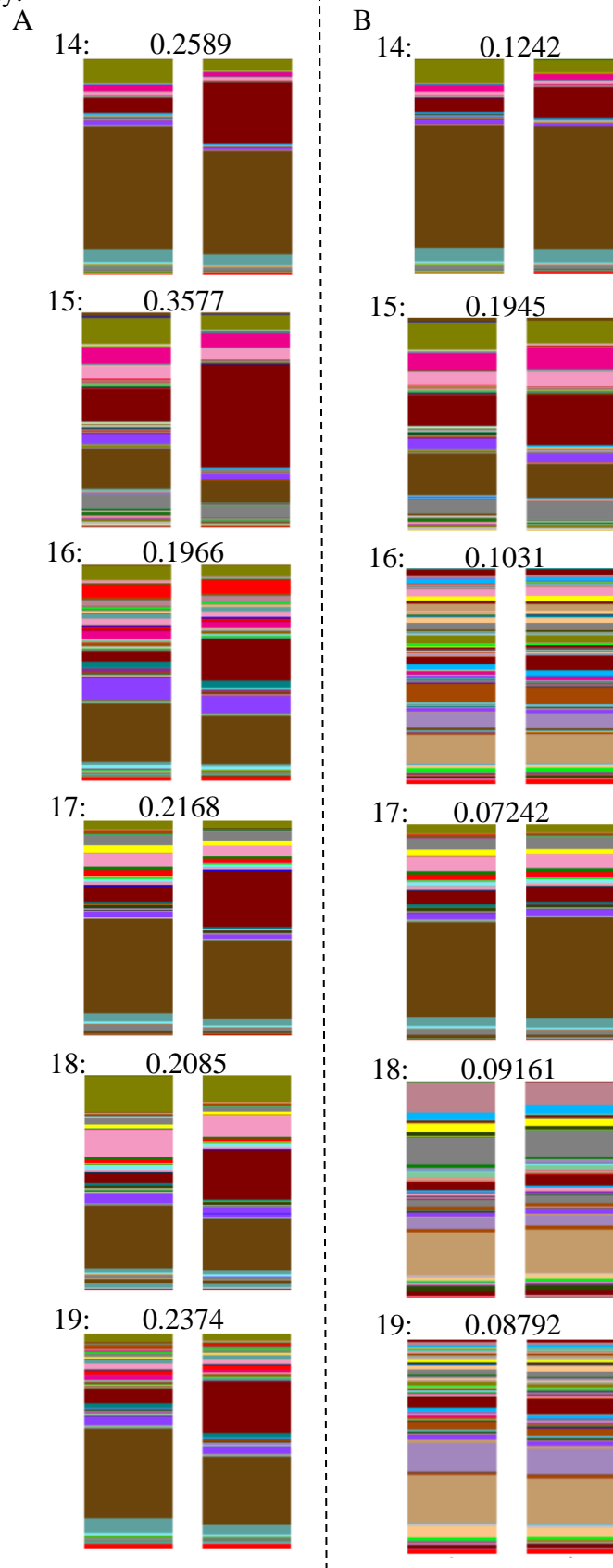


Table 4. Microbial taxa at lower relative abundance in seawater samples when sequenced via the EMP-PNA method versus EMP method. The first column lists rank order abundance of each taxa in the entire seawater sample set. Reported p-values are from paired t-tests with and without false discovery rate correction. Values in () are p-values from Wilcoxon sign-rank tests.

Abund.	P-value	FDR	Taxonomic Classification
# 2	3.12 ⁻⁰⁸ (1.19 ⁻⁰⁷)	2.21 ⁻⁰⁴ (2.39 ⁻⁰⁵)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_
# 11	1.25 ⁻⁰⁶ (1.19 ⁻⁰⁷)	1.54 ⁻⁰³ (2.39 ⁻⁰⁵)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_Octadecabacter
# 50	3.78 ⁻⁰⁶ (1.19 ⁻⁰⁷)	1.66 ⁻⁰⁵ (2.39 ⁻⁰⁵)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_Pseudoruegeria
# 79	1.76 ⁻⁰⁵ (1.19 ⁻⁰⁷)	5.01 ⁻⁰⁴ (2.39 ⁻⁰⁵)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;Other
# 393	1.82 ⁻⁰⁴ (1.43 ⁻⁰⁴)	2.57 ⁻⁰³ (7.92 ⁻⁰³)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Phyllobacteriaceae;Other
# 112	3.75 ⁻⁰⁴ (6.41 ⁻⁰⁵)	2.63 ⁻⁰³ (4.28 ⁻⁰³)	Proteobacteria;c_Alphaproteobacteria;o_Kiloniellales;f_Kiloniellaceae;g_
# 699	6.71 ⁻⁰⁴ (1.66 ⁻⁰³)	2.76 ⁻⁰³ (0.571)	Proteobacteria;c_Gammaproteobacteria;o_34P16;f_;
# 66	0.0010 (6.56 ⁻⁰⁶)	3.49 ⁻⁰³ (5.64 ⁻⁰⁴)	Proteobacteria;c_Alphaproteobacteria;o_BD7-3;f_;
# 187	0.0013 (7.42 ⁻⁰⁴)	7.66 ⁻⁰⁴ (0.030)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Hyphomicrobiaceae;g_
# 175	0.0016 (2.14 ⁻⁰⁴)	1.98 ⁻⁰⁴ (9.91 ⁻⁰³)	Actinobacteria;c_Acidimicrobia;o_Acidimicrobiales;f_TK06;g_
# 83	0.0019 (3.93 ⁻⁰⁶)	7.90 ⁻⁰³ (3.64 ⁻⁰⁴)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_Loktanelia
# 457	0.0033 (2.53 ⁻⁰³)	8.07 ⁻⁰³ (0.072)	Proteobacteria;c_Alphaproteobacteria;o_Rhodobacterales;f_Rhodobacteraceae;g_Sulfitobacter
# 70	0.0036 (1.58 ⁻⁰³)	1.07 ⁻⁰² (0.0559)	Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Phyllobacteriaceae;g_
# 167	0.0039 (4.28 ⁻⁰³)	1.32 ⁻⁰² (0.116)	Planctomycetes;c_OM190;o_CL500-15;f_;
# 478	0.0074 (5.92 ⁻⁰³)	1.34 ⁻⁰² (0.134)	Proteobacteria;c_Gammaproteobacteria;o_Thiotrichales;f_Piscirickettsiaceae;g_Methylophaga
# 759	0.0078 (0.0143)	1.44 ⁻⁰² (0.242)	Chlamydiae;c_Chlamydia;o_Chlamydiales;f_Parachlamydiaceae;Other
# 286	0.0082 (0.007)	1.71 ⁻⁰² (0.153)	Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyrimonadaceae;g_Paludibacter
# 265	0.0090 (0.001)	1.73 ⁻⁰² (0.050)	Proteobacteria;c_Gammaproteobacteria;o_Alteromonadales;f_Alteromonadaceae;g_nsmvVII8

Table 5. Weighted Unifrac distances between replicate samples amplified via the EMP method versus EMP-PNA method show increasing community similarity as the OTUs containing 14-mers and 12-mers matching the pPNA clamp are filtered out.

Sample	Weighted-Unifrac	Weighted Unifrac (no 14-mers)	Weighted Unifrac (no 12-mers)
107 (Aquatic Leaf)	0.04975	0.04709	0.04975
108 (Aquatic Leaf)	0.06905	0.07192	0.06814
109 (Aquatic Leaf)	0.06657	0.06730	0.06366
110 (Aquatic Leaf)	0.1139	0.08982	0.1000
55 (Aquatic Leaf)	0.05650	0.05077	0.05384
56 (Aquatic Leaf)	0.05096	0.04563	0.04806
57 (Aquatic Leaf)	0.04969	0.05072	0.04959
58 (Aquatic Leaf)	0.08909	0.06926	0.07777
1501 (Seawater)	0.1588	0.1137	0.1302
1502 (Seawater)	0.1666	0.08663	0.1217
1503 (Seawater)	0.2752	0.1862	0.2308
1504 (Seawater)	0.2603	0.1975	0.2269
1505 (Seawater)	0.1340	0.07043	0.09493
1506 (Seawater)	0.1959	0.07215	0.1396
1507 (Seawater)	0.1499	0.05116	0.08416
1508 (Seawater)	0.1734	0.06503	0.07265
1509 (Seawater)	0.1633	0.09524	0.1249
1510 (Seawater)	0.1380	0.09350	0.1295
1511 (Seawater)	0.1297	0.08003	0.09384
1512 (Seawater)	0.07023	0.06809	0.07306
1513 (Seawater)	0.08820	0.07787	0.07265
1514 (Seawater)	0.2609	0.1289	0.1676
1515 (Seawater)	0.3561	0.1953	0.2485
1516 (Seawater)	0.1968	0.1003	0.1460
1517 (Seawater)	0.2108	0.06363	0.1586
1518 (Seawater)	0.2168	0.09361	0.1472
1519 (Seawater)	0.2353	0.09416	0.18203
1520 (Seawater)	0.1698	0.1053	0.1280
1521 (Seawater)	0.1747	0.06019	0.1247
1522 (Seawater)	0.1816	0.1234	0.1472
1523 (Seawater)	0.1853	0.07976	0.1416
1524 (Seawater)	0.08419	0.06962	0.06392
253 (Freshwater)	0.07707	0.08273	0.07789
254 (Freshwater)	0.07719	0.07511	0.07295
255 (Freshwater)	0.06102	0.5261	0.05508
256 (Freshwater)	0.04834	0.04735	0.04366
11 (Soil)	0.1823	0.1818	0.1803
12 (Soil)	0.1657	0.1631	0.1633
13 (Soil)	0.1648	0.1641	0.1644
14 (Soil)	0.1663	0.1634	0.1652
15 (Soil)	0.1129	0.1079	0.1099
51 (Terrestrial Leaf)	0.5658	0.5679	0.5690
55 (Terrestrial Leaf)	0.6494	0.7550	0.7164
59 (Terrestrial Leaf)	0.7571	0.7763	0.7802
63 (Terrestrial Leaf)	0.4340	0.4338	0.4353

Appendix S5.

Table 1. Datasets scanned from the Earth Microbiome Project database. Samples were first filtered for low sequence number. Samples with at least 5000 sequences were filtered again for only those taxa listed in Appendix S1 that contain a 14 of 17 bp match to the pPNA clamp. The last column lists the number of samples in each of these datasets that contain at least 1% of these taxa when amplified with EMP primers, suggesting that these types of environmental samples may lead to biased results if sequenced with pPNA clamps. See Table 1 of the main text for a summary of the datasets containing samples in the last column.

Study #	Study Name	Total Samples	Samples w/ at least 5000 sequences	Samples w/ taxa in Appendix S1	# Samples in Table 1
94	Soil bacterial and fungal communities across a pH gradient in an arable soil.	26	0	N/A	N/A
103	Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale.	89	0	N/A	N/A
104	Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes.	52	0	N/A	N/A
213	Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil.	48	1	1	0
214	Microbial consumption and production of volatile organic compounds at the soil-litter interface.	12	0	N/A	N/A
231	Preliminary study of barn swallow microbiome.	83	0	N/A	N/A
314	Characterization of airborne microbial communities at a high-elevation site and their potential to act as atmospheric ice nuclei.	11	0	N/A	N/A
316	Population genetic structure of the prairie dog flea and plague vector, <i>Oropsylla hirsute</i> .	251	0	N/A	N/A
353	Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples.	144	0	N/A	N/A
391	Postprandial remodeling of the gut microbiota in Burmese pythons.	130	0	N/A	N/A
396	The ecology of the phyllosphere: geographic and phylogenetic variability in the distribution of bacteria on tree leaves.	107	1	1	0
619	Neon soils.	337	0	N/A	N/A
632	Canadian metamicrobiome initiative.	13	11	11	3
638	Protist diversity in a permanently ice-covered Antarctic Lake during the polar night transition.	89	89	89	58
659	New Zealand free air carbon dioxide enrichment, agrosearch.	24	23	23	7
662	Intertidal microbes 16S for 2009 and 2010	46	46	46	42
678	Bioturbating shrimp alter the structure and diversity of bacterial communities in coastal marine sediments.	275	257	257	204
713	Diversity of carbonate deposits and basement rocks in continental and marine serpentine seeps.	51	1	1	0
723	Catlin arctic survey 2010 I3.	97	84	84	64
776	Jurelivicius Antarctic cleanup.	30	29	29	2
804	Brazelton LostCity chimney biofilm.	93	78	74	56
805	Effect of soil pH on soil metagenome.	14	14	14	8
807	Gibbons tongue river 16S.	44	44	44	43
808	NEON soils EMP Pilot.	15	13	13	11
809	NEON soils EMP Pilot.	21	19	19	13
810	Ocean Drilling Program Leg 201.	7	3	2	0
829	Environmental metagenomic interrogation of Thar desert microbial communities.	2	2	2	2

846	Influence of tillage practices on soil microbial diversity and activity in a long-term corn experimental field under continuous maize production.	48	48	48	13
861	Comparison of groundwater samples from karst sinkholes (cenotes) from the Yucatan Peninsula, Mexico.	21	21	20	8
864	Magnificent Mongolian microbes.	230	228	228	48
889	Rees Volcano island MedSea.	8	8	8	7
894	Catchment sources of microbes.	1994	1920	1655	375
905	Hulth Gullmarsfjord sediments.	52	52	52	38
910	Viral communities associated with algal/coral interactions.	59	56	24	1
925	Yellowstone gradients.	412	356	136	32
926	Seasonal restructuring of the ground squirrel gut microbiota over the annual hibernation cycle.	46	0	N/A	N/A
929	Bacterial communities associated with the lichen symbiosis.	16	0	N/A	N/A
933	Latitudinal surveys of algal-associated microorganisms.	335	321	321	321
940	Song Colorado freshwater fish.	275	209	174	32
945	Routine samples of German Lakes.	1142	1089	1012	320
963	Green Iguana hindgut microbiome.	100	90	82	6
990	Fermilab spatial study.	708	697	697	29
1001	Cannabis soil microbiome.	26	23	23	20
1024	The soil microbiome influences grapevine-associated microbiota. Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest.	348	100	96	36
1030	Myrold alder fir.	150	147	147	123
1031	Myrold alder fir.	12	11	11	3
1033	Myrold alder fir.	12	5	5	3
1034	CryoCARB permafrost soil microbiome.	90	90	89	45
1035	New Zealand Terrestrial Antarctic Biocomplexity survey (NZTABS).	121	117	117	88
1036	Geochemical landscapes.	68	66	66	14
1037	Long term soil productivity project.	24	24	24	19
1038	Myrold Oregon transect.	21	21	21	14
1039	Rio de Janeiro coastline.	25	23	23	8
1041	Great Lake microbiome.	49	49	49	43
1043	Laboratory directed research and development biological carbon sequestration.	56	54	54	6
1056	Comparison of microbial flora in ant-eating mammals.	93	92	79	14
1064	Bee microbiome.	387	271	72	4
1197	Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill.	106	103	103	101
1198	Polluted polar coastal sediments.	61	57	57	57
1222	Bergen Ocean acidification mesocosms.	72	71	71	71
1235	EPOCA Svalbard 2010.	268	258	258	256
1240	L4 Time Series 2009-2010.	145	140	140	140
1242	Mendota Lake Eleven year time series.	96	91	90	11
1288	Temperate bog lakes.	1505	1350	1342	397
1289	Temple TX native exotic precipitation study.	65	64	64	49
1364	Temporal dynamics in bacterial community of hydra polyyps after hatching	39	3	2	0
1453	Metcalf San Diego Zoo folivorus primate.	316	292	133	0
1485	Predator-prey interactions 18S.	60	58	0	N/A
1526	Recovery of biological soil crust-like microbial communities in previously submerged soils of Glen canyon.	95	95	95	82
1530	Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest.	98	94	94	85
1552	Lake microbial communities are resilient after a whole-ecosystem disturbance.	18	0	N/A	N/A
1578	Ice wedge polygon.	35	35	33	7
1579	Hawaii Kohana volcanic soils.	128	125	117	43

1580	Saline environments that may harbor novel lignocellulolytic activities tolerant of ionic liquids.	26	25	23	8
1621	Saline environments that may harbor novel lignocellulolytic activities tolerant of ionic liquids.	192	188	153	0
1622	Biodiversity and functional patterns of microbial assemblages in postglacial pond sediment profiles.	353	345	245	35
1627	Chu Tibetan plateau lake sediments.	18	18	18	6
1632	Bird egg shells from Spain.	604	527	278	37
1642	Microbial community of the bulk soil and rhizosphere of rice plants over its lifecycle.	644	623	623	25
1665	Marine mammal skin microbiomes.	186	114	86	30
1671	Bacterial communities associated with the surfaces of fresh fruits and vegetables.	214	0	N/A	N/A
1673	Mission Bay sediment viromes.	26	22	14	4
1674	Green roofs as biodiversity corridors in New York City.	151	146	146	135
1692	Friedman Alaska peat soils.	89	75	75	26
1694	Peralta starlings.	562	443	339	114
1696	Comparison of gut flora foliverous primates.	160	157	136	0
1702	Chu Changbai mountain soil.	22	22	22	17
1711	McGuire Kakamenga Kenya soils.	77	71	71	51
1713	Malaysia Lambir Soils.	34	34	34	10
1714	Malaysia Pasoh Landuse logged forest.	25	23	23	10
1715	McGuire Nicaragua coffee soil.	61	60	60	18
1716	Panama precipitation grad soil.	43	41	41	4
1717	McGuire SW Kenya soils.	56	54	54	47
1721	Thomas soil agricultural enhancement.	292	260	246	174
1734	Gut microbiota of Phyllostomid bats that span a breadth of diets.	94	63	39	8
1736	Ezenwa Cape Buffalo.	614	500	468	1
1740	The global sponge microbiome: diversity and structure of symbiont communities across the phylum Porifera.	1403	1206	1098	282
1747	Development of the oral microbiota in captive Komodo dragons (Varanus komodoensis)	210	178	166	22
1773	Garcia bird gut microbiome	122	116	116	76
1792	Diversity and heritability of the maize rhizosphere microbiome under field conditions.	463	213	212	63
1818	Florida decay wastewater study.	198	186	167	52
1845	Variation in the microbiota of Ixodes ticks with geography, species and sex Illumina.	124	91	56	8
1883	Crump Arctic LTREB main.	3153	2415	2368	794
1885	Variation in the Microbiota of Ixodes ticks with geography, species and sex.	139	16	10	0
2019	Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate.	272	81	60	0
2020	Study 2020.	98	0	N/A	N/A
2080	Seyler North Atlantic water column.	54	53	53	26
2104	Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally 16S.	1160	1160	1160	632
2182	Hale folivorous primates.	167	162	78	4
2229	Thomas CMB Australian seaweed.	1378	1285	1285	1270
2259	Individuals diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch).	62	46	31	5
2300	Gut microbiome of hibernating bears.	96	68	16	0
2338	Song whitehead bats.	192	102	30	6
2382	The soil microbiome influences grapevine-associated microbiota HiSeq.	401	315	309	106
10119	Microbial biogeography of grapes predicts regional metabolite patterns in wine.	47	0	N/A	N/A
10141	Metcalf microbial community assembly and metabolic function during mammalian corpse decomposition mouse exp.	68	0	N/A	N/A

10142	Metcalf microbial community assembly and metabolic function during mammalian corpse decomposition SHSU winter.	104	0	N/A	N/A
10143	Metcalf microbial community assembly and metabolic function during mammalian corpse decomposition SHSU April 2012 exp.	927	796	0	N/A
10145	Beach sand microbiome from Calvert Island Canada.	114	91	91	86
10156	The effect of wetland age and restoration methodology on long term development and ecosystem functions of restored wetlands.	192	179	178	47
10180	Metagenome of microbial communities involved in the nitrogen cycle in sugarcane soils in Brazil.	128	112	110	36
10196	Composition of symbiotic bacteria as a predictor of survival in Panamanian golden frogs infected with <i>Batrachochytrium dendrobatidis</i> .	37	37	36	2
10245	Diversity, host affinity and ecology of foliar endophytic microbes in Amazonian Peru.	120	103	61	7
10246	The North American Arctic Transect, NAAT and the Eurasian Arctic Transect, EAT.	112	70	70	58
10272	Most of the dominant members of amphibian skin bacterial communities can be readily cultured.	64	64	59	31
10273	SM April WHOI SeaWater	67	45	45	23
10278	Identifying the microbial cohorts associated with drought-driven carbon release from peatland ecosystems.	216	215	215	29
10308	Whitehead fish.	1208	938	697	172
10311	Ecological succession reveals signatures of marine–terrestrial transition in salt marsh fungal communities.	58	0	N/A	N/A
10324	Diversity of Rickettsiales in the microbiome of the Lone Star Tick, <i>Amblyomma americanum</i> .	87	1	1	1
10346	The global sponge microbiome: diversity and structure of symbiont communities across the phylum Porifera – final.	1390	1194	1068	285
10363	Investigating the rhizosphere microbiome as influenced by soil selenium, plant species, plant selenium accumulation and geographic proximity.	64	58	58	55
10369	Obligate biotroph pathogens defend their niche against competing microbes by keeping host defense at a functional level.	9	0	N/A	N/A
10376	Muegge mammals.	22	22	17	0

Appendix S6.

Table 1. The 97% OTU Greengenes database (version 13_8, containing 99,322 sequences) was scanned for matches to 12-mer through 17-mer combinations, including gaps, of the pPNA chloroplast blocking clamp and the mPNA mitochondrial blocking clamp. We note that the Greengenes database contains 67 sequences identified as chloroplast, and so we list this total number of hits in parentheses, however the bolded number in the Matches column equals the number of bacteria OTU hits excluding these organelle sequences. Corresponding file names are pPNA for chloroplast clamps and mPNA for mitochondrial clamps.

pPNA		Matches	Filenames	mPNA	
pPNA n-mers				mPNA n-mers	Matches
17mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA17mer.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
16mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA16merA.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
16mer: <i>GGCTCAACCCTGGACAG</i>	0 (60)		_PNA16merB.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
15mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA15merA.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
15mer: <i>GGCTCAACCCTGGACAG</i>	0 (60)		_PNA15merB.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
15mer: <i>GGCTCAACCCTGGACAG</i>	60 (124)		_PNA15merC.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
14mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA14merA.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
14mer: <i>GGCTCAACCCTGGACAG</i>	0 (60)		_PNA14merB.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
14mer: <i>GGCTCAACCCTGGACAG</i>	60 (124)		_PNA14merC.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
14mer: <i>GGCTCAACCCTGGACAG</i>	1,338 (1,405)		_PNA14merD.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
13mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA13merA.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
13mer: <i>GGCTCAACCCTGGACAG</i>	0 (60)		_PNA13merB.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
13mer: <i>GGCTCAACCCTGGACAG</i>	60 (124)		_PNA13merC.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
13mer: <i>GGCTCAACCCTGGACAG</i>	0 (59)		_PNA13merD.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
13mer: <i>GGCTCAACCCTGGACAG</i>	1,820 (1,887)		_PNA13merE.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	12 (71)		_PNA12merA.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	2 (62)		_PNA12merB.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	60 (124)		_PNA12merC.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	1,344 (1,411)		_PNA12merD.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	1,826 (1,893)		_PNA12merE.fna	<i>GGCAAGTGTTCCTCGGA</i>	0 (38)
12mer: <i>GGCTCAACCCTGGACAG</i>	2,314 (2,381)		_PNA12merF.fna	<i>GGCAAGTGTTCCTCGGA</i>	2 (40)

Table 2. The 97% Silva database (version 123, containing 226,267 sequences) was scanned for matches to 12-mer through 17-mer combinations of the pPNA chloroplast blocking clamp and the mPNA mitochondrial blocking clamp. We note that the Silva database contains 1689 sequences identified as chloroplast and 529 sequences identified as mitochondria, and so we list this total number of hits in parentheses, however the bolded number in the Matches column equals the number of bacteria OTU hits excluding these organelle sequences. Corresponding file names are pPNA for chloroplast clamps and mPNA for mitochondrial clamps.

pPNA			mPNA		
pPNA n-mers	Matches	Filenames	mPNA n-mers	Matches	
17mer: <i>GGCTCAACCCTGGACAG</i>	333	_PNA17mer.fna	<i>GGCAAGTGTCTTCGGA</i>	190	
16mer: <i>GGCTCAACCCTGGACAG</i>	339	_PNA16merA.fna	<i>GGCAAGTGTCTTCGGA</i>	191	
16mer: <i>GGCTCAACCCTGGACAG</i>	353	_PNA16merB.fna	<i>GGCAAGTGTCTTCGGA</i>	192	
15mer: <i>GGCTCAACCCTGGACAG</i>	344	_PNA15merA.fna	<i>GGCAAGTGTCTTCGGA</i>	193	
15mer: <i>GGCTCAACCCTGGACAG</i>	359	_PNA15merB.fna	<i>GGCAAGTGTCTTCGGA</i>	193	
15mer: <i>GGCTCAACCCTGGACAG</i>	468	_PNA15merC.fna	<i>GGCAAGTGTCTTCGGA</i>	199	
14mer: <i>GGCTCAACCCTGGACAG</i>	348	_PNA14merA.fna	<i>GGCAAGTGTCTTCGGA</i>	199	
14mer: <i>GGCTCAACCCTGGACAG</i>	364	_PNA14merB.fna	<i>GGCAAGTGTCTTCGGA</i>	195	
14mer: <i>GGCTCAACCCTGGACAG</i>	474	_PNA14merC.fna	<i>GGCAAGTGTCTTCGGA</i>	200	
14mer: <i>GGCTCAACCCTGGACAG</i>	3,235	_PNA14merD.fna	<i>GGCAAGTGTCTTCGGA</i>	204	
13mer: <i>GGCTCAACCCTGGACAG</i>	350	_PNA13merA.fna	<i>GGCAAGTGTCTTCGGA</i>	214	
13mer: <i>GGCTCAACCCTGGACAG</i>	371	_PNA13merB.fna	<i>GGCAAGTGTCTTCGGA</i>	201	
13mer: <i>GGCTCAACCCTGGACAG</i>	479	_PNA13merC.fna	<i>GGCAAGTGTCTTCGGA</i>	203	
13mer: <i>GGCTCAACCCTGGACAG</i>	3,246	_PNA13merD.fna	<i>GGCAAGTGTCTTCGGA</i>	205	
13mer: <i>GGCTCAACCCTGGACAG</i>	4,258	_PNA13merE.fna	<i>GGCAAGTGTCTTCGGA</i>	210	
12mer: <i>GGCTCAACCCTGGACAG</i>	369	_PNA12merA.fna	<i>GGCAAGTGTCTTCGGA</i>	242	
12mer: <i>GGCTCAACCCTGGACAG</i>	374	_PNA12merB.fna	<i>GGCAAGTGTCTTCGGA</i>	220	
12mer: <i>GGCTCAACCCTGGACAG</i>	486	_PNA12merC.fna	<i>GGCAAGTGTCTTCGGA</i>	210	
12mer: <i>GGCTCAACCCTGGACAG</i>	3,259	_PNA12merD.fna	<i>GGCAAGTGTCTTCGGA</i>	213	
12mer: <i>GGCTCAACCCTGGACAG</i>	4,274	_PNA12merE.fna	<i>GGCAAGTGTCTTCGGA</i>	211	
12mer: <i>GGCTCAACCCTGGACAG</i>	5,308	_PNA12merF.fna	<i>GGCAAGTGTCTTCGGA</i>	215	