

Supplementary Table 6

Supplementary Table 6: Comparison of *PhyloPythia* to the SOM-phyloptype associations and tetranucleotide-based binning of the dominant sample populations for the contigs ≥ 1 kb of the Sargasso Sea sample. The upper part gives the respective values for the high-coverage contigs only; the lower part for the complete set of all contigs ≥ 1 kb. The 25 phylotype groupings of the SOM are non-overlapping groups but at different taxonomic ranks of NCBI Taxonomy. For comparison with our technique, we give the results at class level for the subset of sequences where it applies, and at phylum level, where for each lower-level clade a parent is defined. '# Contigs' gives the number of contigs that are part of each clade in the reference. The notation ' x (y)' denotes for each of the clades the number of correctly assigned fragments x and the number of false positive assignments y . 'A' denotes the percentage of correctly assigned items (note that unassigned items are counted 'false'), 'Asg.' denotes the percentage of assigned items and *Sp.* the specificity of assignments. For the higher-level *Sp.*-values, assignments are counted as 'false' that are not assigned to a node at a level, but are annotated with a parent node that disagrees with the current level assignments (e.g. Proteobacteria as an assignment for an archaeal node).

Dominant sample pops.	<i>A</i> _{contig} (%)	<i>Asg</i> _{contig} (%)	<i>Sp</i> _{contig} (%)	<i>A</i> _{nt} (%)	<i>Asg</i> _{nt} (%)	<i>Sp</i> _{nt} (%)
# Contigs _{SSU:rRNA-highcoveragecontigs}	46			1,902,990		
PhyloPythia _{ab}	72	72	100	98.21	98.21	100
TETRA*	39	39	100	93.87	93.87	100
Class-level						
# Contigs _{SSU:rRNA-highcoveragecontigs}	46			1,902,990		
PhyloPythia _c	74	78	94	98.94	99.08	99.86
SOM	20	19	100	49.21	c	100

Dominant sample pops.						
# Contigs _{SSU:rRNA\geq1kcontigs}	69			1,994,724		
PhyloPythia _{ab}	53.6	53.6	100	96.36	96.36	100
TETRA*	27.5	27.5	100	90.79	90.79	100
Class-level						
# Contigs _{SSU:rRNA\geq1kcontigs}	333			2,091,355		
PhyloPythia _c	18	27.3	62.5	75.89	80.3	93.48
SOM	4.2	6.6	60.9	35.99	36.7	97.86
Phylum-level						
# Contigs _{SSU:rRNA\geq1kcontigs}	393			2,632,674		
PhyloPythia _c	26.7	52.2	48	76.64	86.63	86.66
SOM	5.3	33.3	15.9	35.02	45.6	76.67

Dominant sample pops.	Burkholderia	Shewanella	New Gammaproteobacteria	Other organisms
# Contigs _{SSU:rRNA-highcoveragecontigs}	8	13	17	32
PhyloPythia _{ab}	7 (0)	11 (0)	13 (0)	32 (-)
TETRA*	7 (0)	7 (0)	4 (0)	32 (-)
Class-level				
	Betaproteobacteria	Gammaproteobacteria		
# Contigs _{SSU:rRNA-highcoveragecontigs}	8	30		
PhyloPythia _c	7 (1)	25 (0)		
SOM	7 (0)	2 (0)		

Dominant sample pops.	Burkholderia	Shewanella	New Gammaproteobacteria				
# Contigs _{SSU:rRNA\geq1kcontigs}	19	23	17				
PhyloPythia _{ab}	8 (0)	14 (0)	13 (0)				
TETRA*	7 (0)	8 (0)	4(0)				
Class-level							
	Betaproteobacteria	Gammaproteobacteria	Alphaproteobacteria	Flavobacteria	Mollicutes	Planctomycetacia	
# Contigs _{SSU:rRNA\geq1kcontigs}	20	190	80	28	1	2	
PhyloPythia _c	16 (0)	41 (15)	1 (9)	0 (5)	0 (0)	0 (0)	
SOM	7 (0)	7 (4)	0 (3)	0 (1)	0 (0)	0 (0)	
Phylum-level							
	Bacteroidetes	Firmicutes	Planctomycetes	Proteobacteria			
# Contigs _{SSU:rRNA\geq1kcontigs}	28	1	2	350			
PhyloPythia _c	0 (12)	0 (0)	0 (0)	99 (86)			
SOM	0 (15)	0 (0)	0 (0)	21 (95)			