

Supplementary Table 7

Supplementary Table 7: Evaluation of *PhyloPythia*'s classification accuracy for genome fragments of different *Prochlorococcus* strains. At the extremes, the average genomic GC content differs by more than 27% (column % GC). Column Accuracy₁ gives the success rate in assigning 50 kb *Prochlorococcus* fragments as *Prochlorococcus* (in all cases as opposed to non-*Prochlorococcus*, built with non-*Prochlorococcus* fragments from the completed genomes) with models built from fragments of opposing GC content (e.g. the assignment of low GC content sequences with the high GC content model). Column Accuracy₂ gives the assignment accuracy in a 'leave-one out' experiment, where the *Prochlorococcus*-model was trained with fragments from all genomes except the one being tested. Column Accuracy₃ gives the assignment accuracy for an all-genome-*Prochlorococcus*-model, which was built from fragments of all of the genomes.

TaxID	Name	% GC	Accuracy ₁ (Opp. GC-models)	Accuracy ₂ (Leave-one-out model)	Accuracy ₃ (All genomes model)
167539	<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375	36.5	97.6	100	100
59919	<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1986	30.8	0.7	100	100
59920	<i>Prochlorococcus marinus</i> str. NATL2A	35.1	72.6	100	100
74546	<i>Prochlorococcus marinus</i> str. MIT 9312	31.2	1.7	100	100
74547	<i>Prochlorococcus marinus</i> str. MIT 9313	50.7	0	0	100
93059	<i>Prochlorococcus marinus</i> str. MIT 9211	58.2	93.1	94	100