

Table S4: Summary of the MgFC classifier results tested on an independent data from [1], using partial fragments

<b>Sequence Length</b>	<b>Number of runs</b>	<b>Fraction of sequences with confirmed results*</b>	<b>Standard deviation</b>
400	25	0.85	0.03
300	30	0.75	0.04
200	50	0.72	0.04
100	100	0.63	0.05

\* Sequences were considered confirmed only when both our classifiers converged and the prediction results fitted the original prediction.

#### Reference

1. Chenard C, Suttle CA: **Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters**. *Appl Environ Microbiol* 2008, **74**(17):5317-5324.