# Supporting Information

Coleman and Chisholm 10.1073/pnas.1009480107

SI Text

The pyrosequencing reads have been deposited in the NCBI Sequence Read Archive, http://www.ncbi.nlm.nih.gov, under the following accession numbers: SRX007372 (HOT186 25m), SRX007369 (HOT186 75m), SRX007370 (HOT186 110m), SRX008032 (BATS216 20m), SRX008033 (BATS216 50m), SRX008035 (BATS216 100m). Shotgun sequence libraries have been deposited in the NCBI Trace Archive with the following trace identifiers: 2243992048-2244135183 (BATS216 20m), 2244135184-2244269775 (BATS216 50m), 2281908320-2281966591 (HOT186 25m), 2281966591-2282003263 and 2282006336-2282042335 (HOT186 75m).
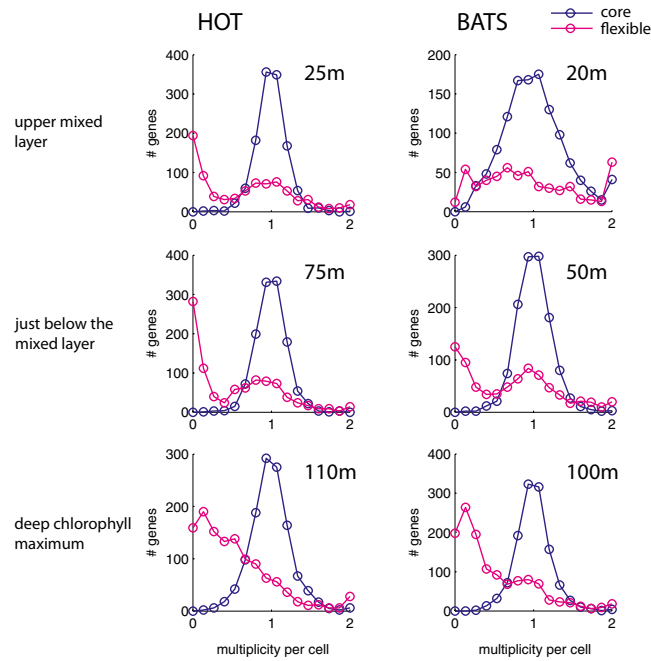
Fig. S1. Copies per *Prochlorococcus* cell of core and flexible genes at each depth. Core genes, defined using whole genome sequences from cultured isolates, also are one copy per cell in these populations, whereas many flexible genes are rare (present in only a subset of cells) at any given depth. Note that the 110-m sample from HOT and the 20-m sample from BATS each had significantly fewer *Prochlorococcus* reads (Table S1), resulting in the unusual shape of the flexible gene distribution.

Fig. S2. Expression of phosphonate utilization genes in *Prochlorococcus* strain MIT9301 under P-starvation conditions. Two control genes—*pstS*, encoding the substrate-binding protein for phosphate, and *rbcL*, encoding the large subunit of RuBisCO—are plotted for comparison. Exponentially growing cells in P-replete media were resuspended just before time 0 in a medium with no added phosphate. For each gene, expression is normalized to that of the control gene, *rnpB*. Error bars show 2 SDs of two biological replicates. The locations of these genes in the genomes are denoted with an asterisk in Fig. 2*B*.

Fig. S3. Gene order variability surrounding core genes and BATS-enriched genes. We identified small-insert clones in which one read matched either a core gene (*A* and *C*) or a BATS-enriched gene (*B* and *D*). For each gene, we asked how many unique COGs are found adjacent to this gene of interest in the clone library. The distribution of COGs adjacent to *Prochlorococcus* BATS-enriched genes (*B*) is significantly different from that of COGs adjacent to *Prochlorococcus* core genes (*A*; χ^2 goodness-of-fit test, $P = 0.002$). The two distributions are not significantly different for *Pelagibacter* (*C* and *D*; $P = 0.6$).

Other Supporting Information Files

Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)
Table S5 (DOC)
Table S6 (DOC)