Bioinformatic Methods

All tools employed in this study a

**RNAseq raw data processing**

Pair-end reads were first:

1. Quality trimming: AfterQC + get only reads from the “good” folder. Using default settings.
2. Remove phiX174
3. Remove rRNA

**Sequence alignment and read counting**

Then pair-end reads were mapped to the Prochlorococcus MED4 genome (accession number: [NC\_005072](https://www.ncbi.nlm.nih.gov/nuccore/NC_005072)) using Bowtie2 [[60](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898218/" \l "B60)] with at most one mismatch. The coverage of each nucleotide was calculated by counting the number of reads mapped at corresponding nucleotide positions in the genome. The number of reads that were perfectly mapped to a gene region was calculated using BEDTools [[61](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898218/" \l "B61)], and then it was normalized by gene length and total mapped reads, namely RPKM as the gene expression value [[26](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898218/" \l "B26)]. The gene annotations for Prochlorococcus MED4 were downloaded from MicrobesOnline [[62](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898218/" \l "B62)] with modifications for non-annotated genes that were designated “HyPMM#”. New ORFs identified in this study were annotated with “TibPMM#” (Sheet 2 of Additional file [3](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898218/" \l "S3)). Sequences generated by this study are available in the Gene Expression Omnibus (GEO) under accession number [GSE49517](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49517).

**Differential expression analysis**

**Gene clustering**

**Gene function annotation**