

Supplemental Table Legends

Table S1. Protein subcellular localization of major marine bacterial taxa using genomic, metagenomic, and metatranscriptomic sequences.

Table S2. GOS samples from six geographic locations and assignment of sequences to taxonomic groups based on best BLAST hit.

Table S3. Performance of the previous version of MetaP in predicting non-membrane proteins.

Table S4. Performance of the new MetaP in predicting non-membrane proteins.

Table S5. Performance of PSLDOC in predicting inner and outer membrane proteins.

Table S6. Performance of CELLO in predicting inner and outer membrane proteins.

Table S7. Predicted subcellular localization of proteins in sequenced genomes of dominant surface ocean bacterioplankton groups

(<http://www.marsci.uga.edu/facultypages/moran/publications/supplements.html>).

Table S8. Predicted subcellular localization of proteins in GOS open ocean and coastal water samples (<http://www.marsci.uga.edu/facultypages/moran/publications/supplements.html>).

Table S9. Predicted subcellular localization of proteins in metatranscriptomics samples (<http://www.marsci.uga.edu/facultypages/moran/publications/supplements.html>).

Table S10. Pearson's correlation coefficient between subcellular localization and axis coordinates in two-dimensional space.

Table S11. Number of gene transcripts in day versus night samples binned to COG functional categories for different subcellular localized proteins.

Table S12. Relative change (%) of COG family abundance (normalized by library size) between day and night gene expression for predicted cytoplasmic, inner membrane, and periplasmic proteins.

Supplemental Data Set

Data Set 1. Protein sequences used as queries in BLAST searches for phototrophy-related genes in metatranscriptomic datasets.

Data Set 2. Protein sequences used to test the accuracy of MetaP software.