Sample Acquisition and Processing
R/V Atlantic Explorer (AE1409)

Trichodesmium sequence accessions: https://www.bco-dmo.org/dataset/716817

Project: Dissolved Phosphorus Processing by Trichodesmium Consortia: Quantitative Partitioning, Role of Microbial Coordination, and Impact on Nitrogen Fixation (P Processing by Tricho)
Principal Investigator: Dr Sonya T. Dyhrman (Lamont-Doherty Earth Observatory, LDEO)
2017-10-13

Sampling:
Trichodesmium colonies were collected with surface water net tows. Sampling occurred at the same time each day using nets with a mesh size of 130 µm. Nets were deployed and hauled through the surface water column six times before recovery, such that each sample represented thousands of liters of water. Individual *Trichodesmium* colonies were isolated and washed three times by successive transfer through fresh 0.2 µm sterile-filtered local surface seawater to remove all but tightly associated epibionts. A pooled sample of colonies was isolated and processed from each station. For each sample, an average of ~30 cleaned colonies were transferred onto 47 mm 5 µm pore size polycarbonate filters, gently vacuum filtered to remove excess liquid, flash frozen and stored in liquid nitrogen until extraction and sequencing.

Processing:
Metagenomic reads were first trimmed using Sickle with default settings (https://github.com/najoshi/sickle). Trimmed forward and reverse reads were then converted to fasta with the fq2fa command in IDBA-UD. Reads from the six samples were assembled into scaffolds to create a merged assembly, using IDBA-UD under default parameters in order to yield robust assembly of the western North Atlantic *Trichodesmium* holobiont. This merged assembly of pooled colony metagenomic reads from across six stations was used for the bulk of the analyses presented herein.

Scaffolds produced by the merged assembly were clustered into genome bins by tetranucleotide frequency and read coverage of individual samples using MaxBin 2.0 set with default parameters. Genome completeness was estimated at >65% using MaxBin, resulting in robust gene set comparisons for the majority of genome bins. Relative abundance estimates were calculated by multiplying the length of contigs in each bin by the number of reads recruited (coverage), then summing across genome bins. Binned scaffolds were translated into predicted proteins using Prodigal on the metagenomic setting. The resulting protein sequences were annotated using the blastp program of DIAMOND against the NCBI nr database.

Metatranscriptome sequence reads were trimmed and normalized following the Eel Pond Protocol for mRNAseq assembly. To obtain read counts for each sample, cleaned forward and reverse reads were mapped to metagenome assemblies from the previously described metagenomes. Mapping was carried out using RSEM using the paired-end and Bowtie2 parameters. Gene read counts were then summed for *Trichodesmium* and epibiont genome bins separately. Significant differential expression of genes between control and +AHL samples was

determined by comparing control and sample treatments using a stringent empirical Bayes approach called Analysis of Sequence Counts (ASC). Epibiont genes were considered significantly differentially expressed between control and +AHL treatment if they had a 95% or higher posterior probability of a fold change greater than 2, as determined by the Analysis of Sequence Counts (ASC). Relative abundance for metatranscriptome experiments was calculated by normalizing read counts to total bin length.