

Appendix A

Supplemental Information

A.1 Appendix for Chapter 2

A.1.1 Supplemental Figures

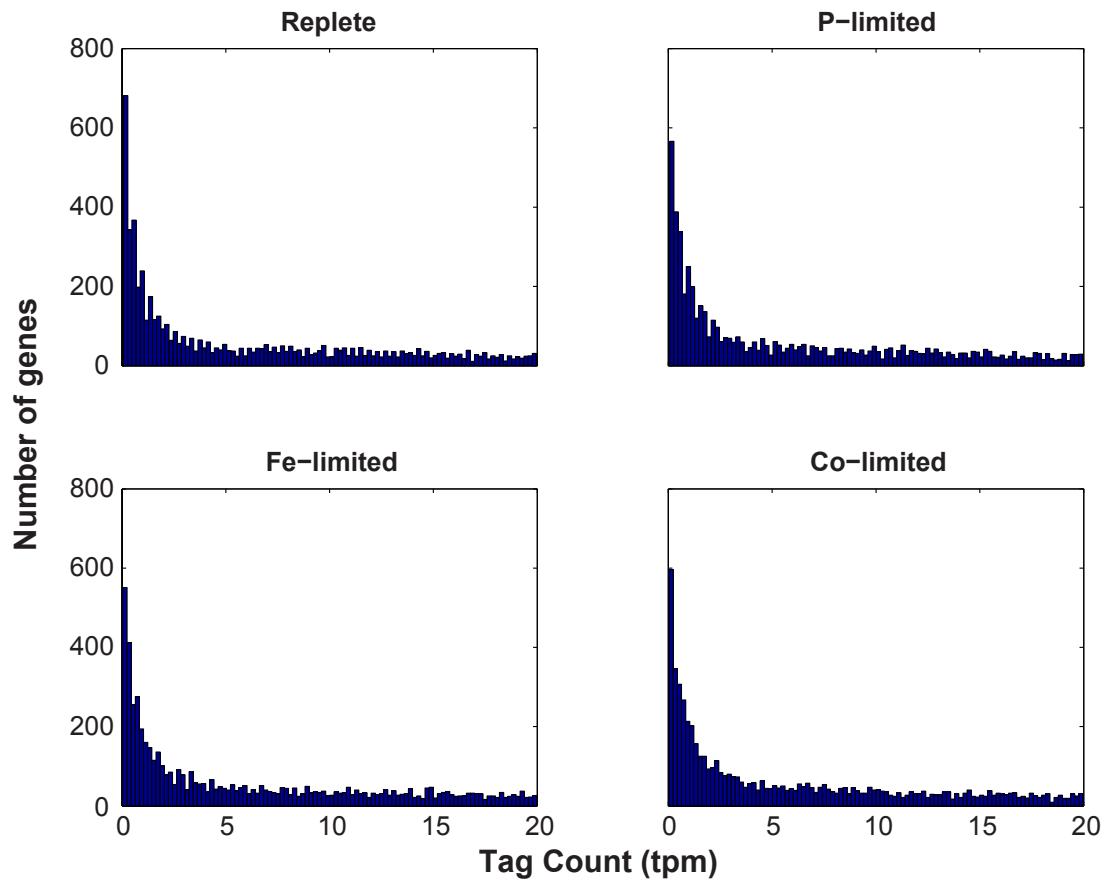


Figure A-1: Histogram analysis of the distribution of normalized tag counts (TPM) for each gene across each of the four treatments (Replete, P-limited, Fe-limited, and co-limited). The abundance of normalized tag counts (TPM) was assessed, tallying the total number of genes with a given tag count. Only tag counts less than 20 are depicted to aid the visualization of the inflection in the data at 2.5 TPM.

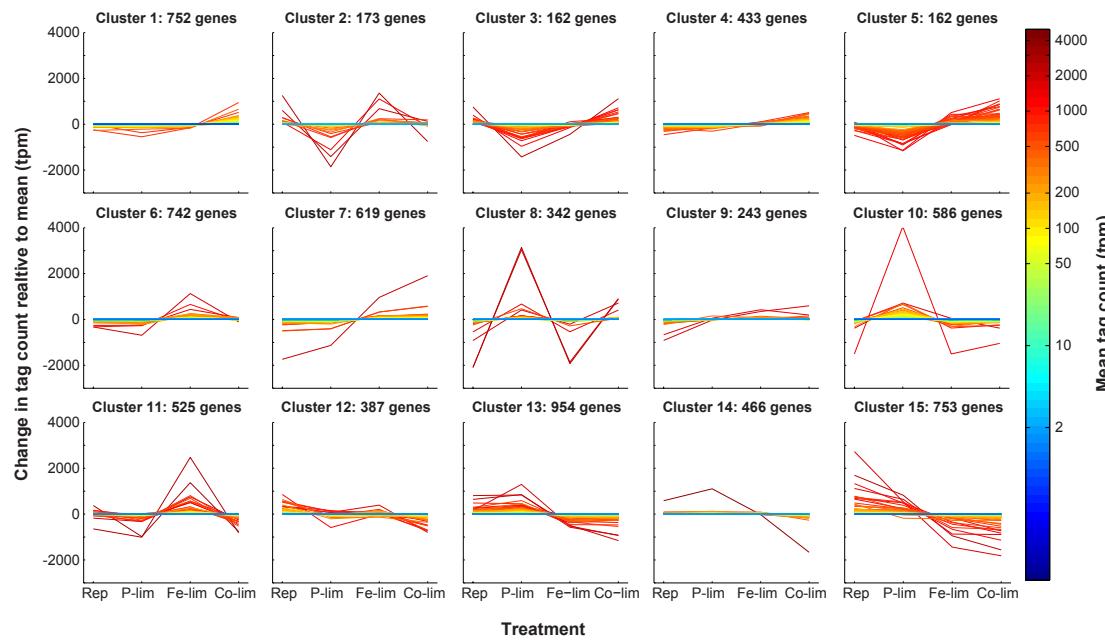


Figure A-2: K -means clustering of normalized genes. The 7380 genes that passed the 2.5 TPM cutoff were clustered into 15 clusters using the k -means algorithm under the Pearson correlation coefficient. Tag counts normalized to total library size (in TPM) for each gene are plotted relative to the mean (indicated by the color of the line) for each of the four treatments: Replete (Rep), P-limited (P-lim), Fe-limited (Fe-lim), and co-limited (Co-lim).

A.2 Appendix for Chapter 3

A.2.1 Supplemental Figures

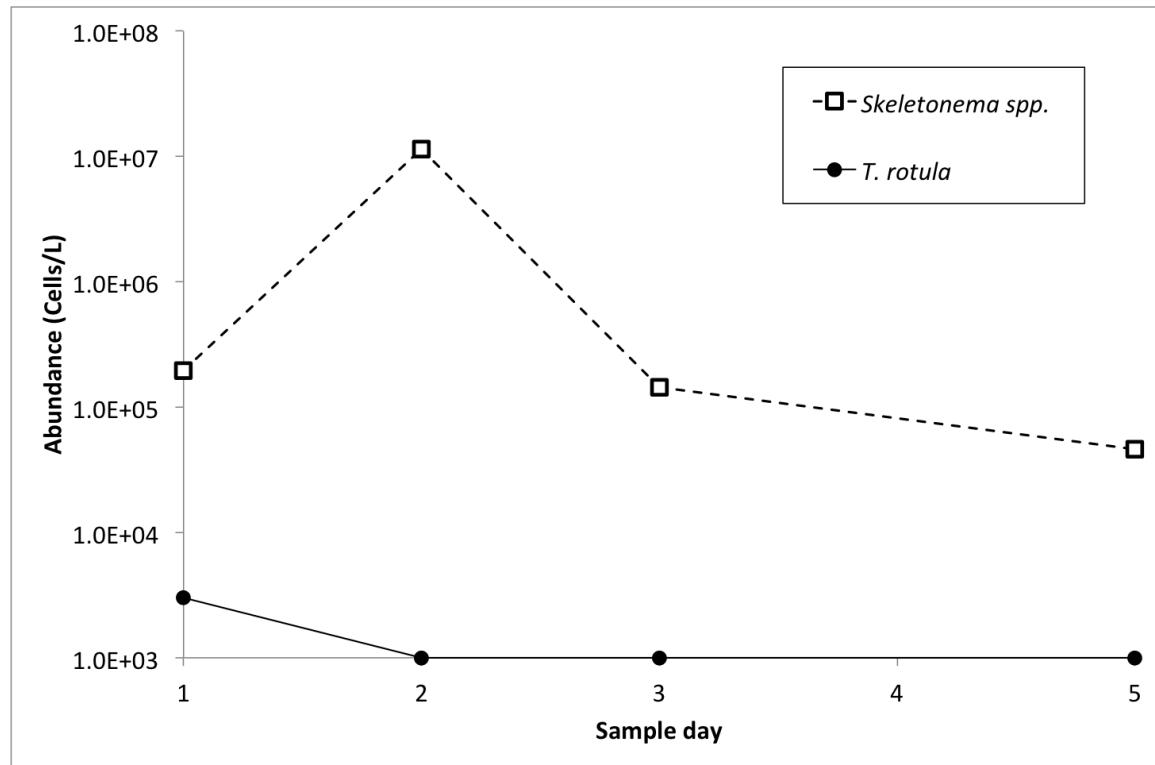


Figure A-3: Abundance estimation from cell counts of *Skeletonema* spp. and *T. rotula* across the five sample points during the spring of 2012.

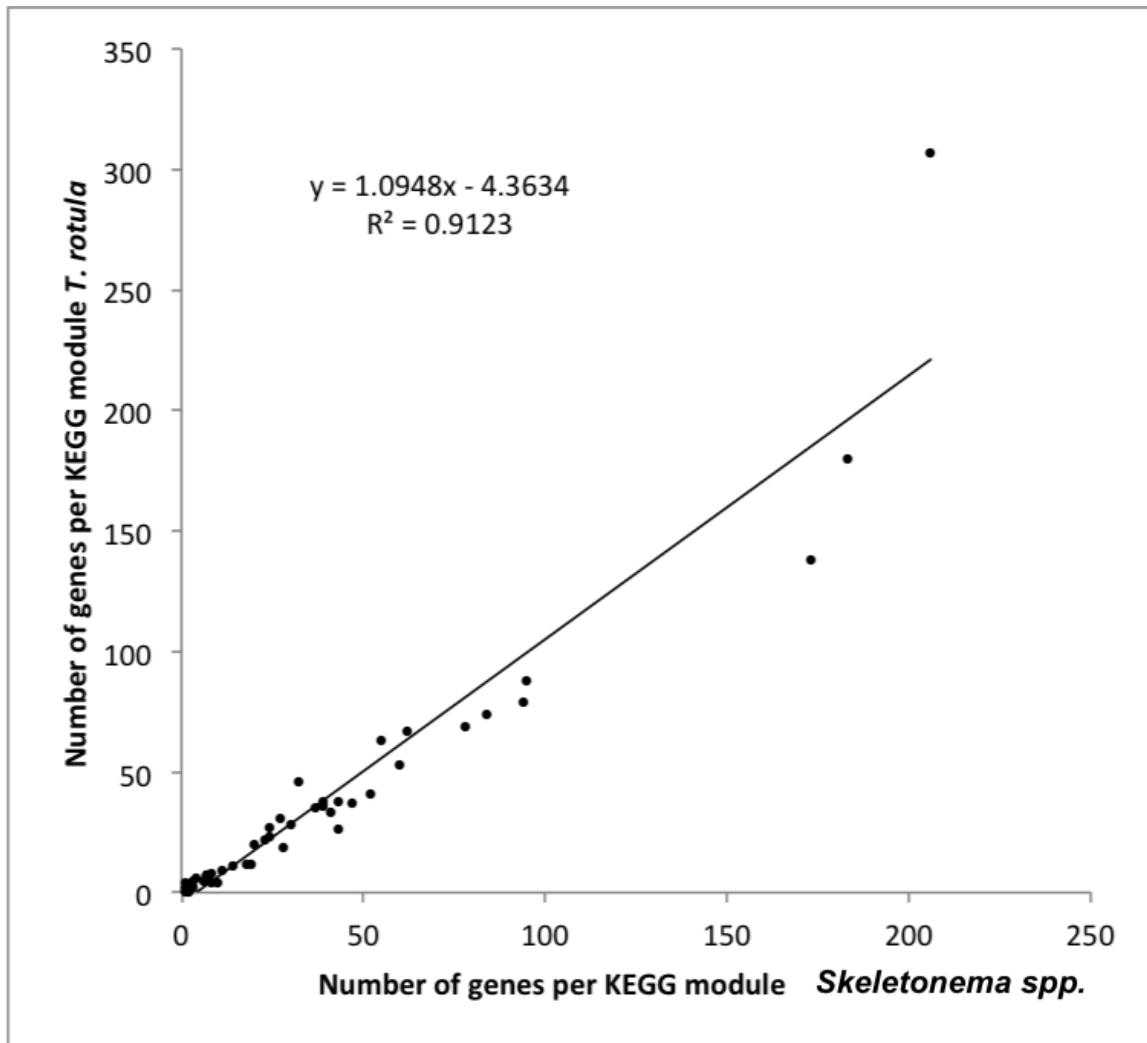


Figure A-4: Total number of genes assigned to each KEGG module for *Skeletonema* spp. and *T. rotula*.

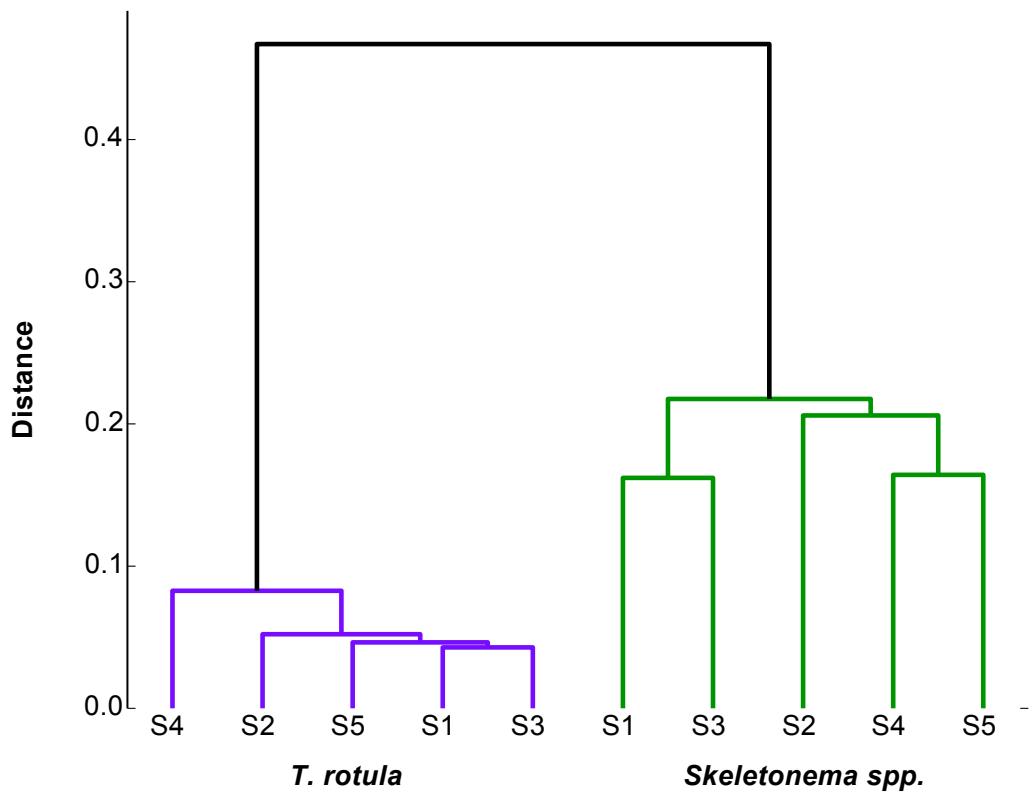


Figure A-5: Dendrogram depicting hierarchical clustering of samples based on relative expression of KEGG modules (Figure 2) across the five samples S1-S5 for *Skeletonema* spp. and *T. rotula*.

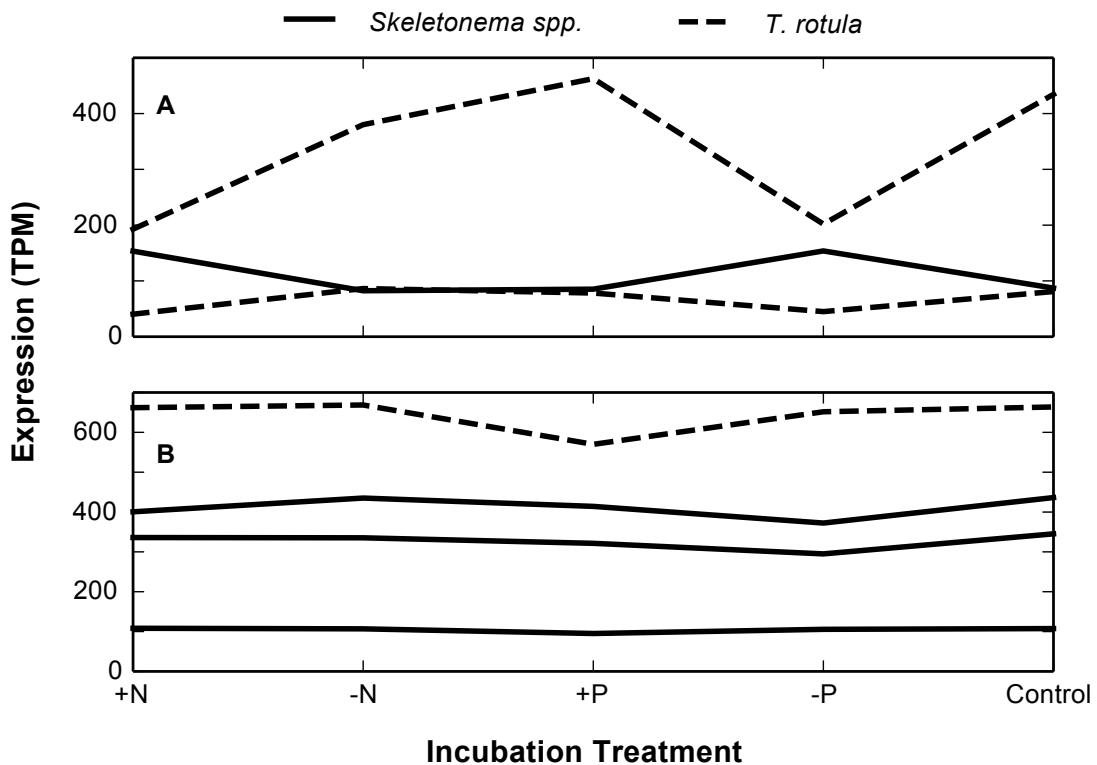


Figure A-6: Expression of stable reference genes identified based on literature and statistical parsing in nutrient amendment incubation. (A) The expression in tags per million (*TPM*) of stable reference genes identified in *T. rotula* (dashed line) and *Skeletonema* spp. (solid lines) based on homology (*e*-value < 1e-5) to a known reference genes in *T. pseudonana*, ACT1 (Thaps_25772), in nutrient incubations. (B) Also shown are reference genes identified in the incubation experiments, using statistical analysis of sequence counts (Alexander et al., 2012; Wu et al., 2010), and nutrient incubations

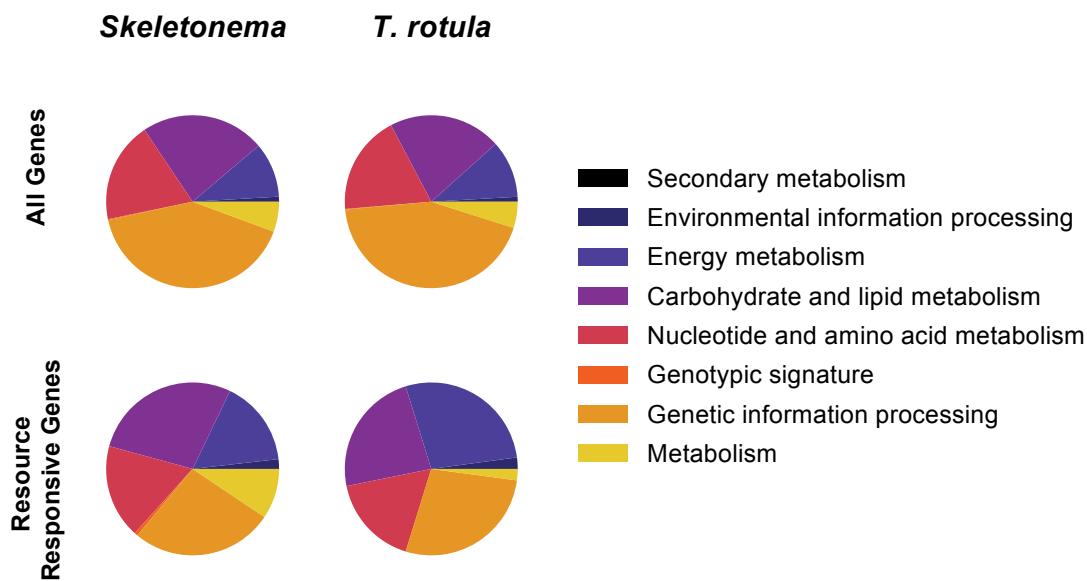


Figure A-7: Functional composition of the reference transcriptome and resource-responsive (RR) gene subset for *T. rotula* and *Skeletonema* spp. (A) RR gene sets were identified through cross comparison of like-nutrient incubations (i.e. +N vs. -N and +P vs. -P), using ASC (fold change = 2, post- $p > 0.95$). The relative functional categorization of the reference transcriptomes and RR gene set for *T. rotula* and *Skeletonema* spp. based on KEGG ontology as assigned by KAAS is depicted at the module-level.

Nitrate Reductase

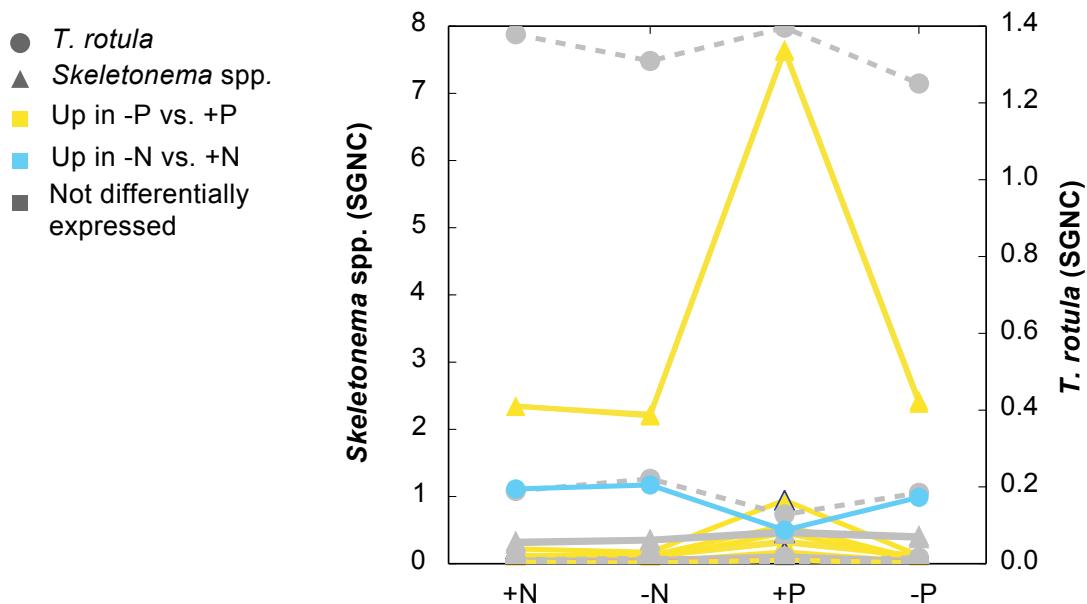


Figure A-8: The relative expression in stable gene normalized counts ($SGNC$) of the assimilatory nitrate reductase gene cluster across the incubation experiment treatments. Significance of regulation between the treatments is denoted by the color of the line; organisms are denoted by the shapes of the marker.

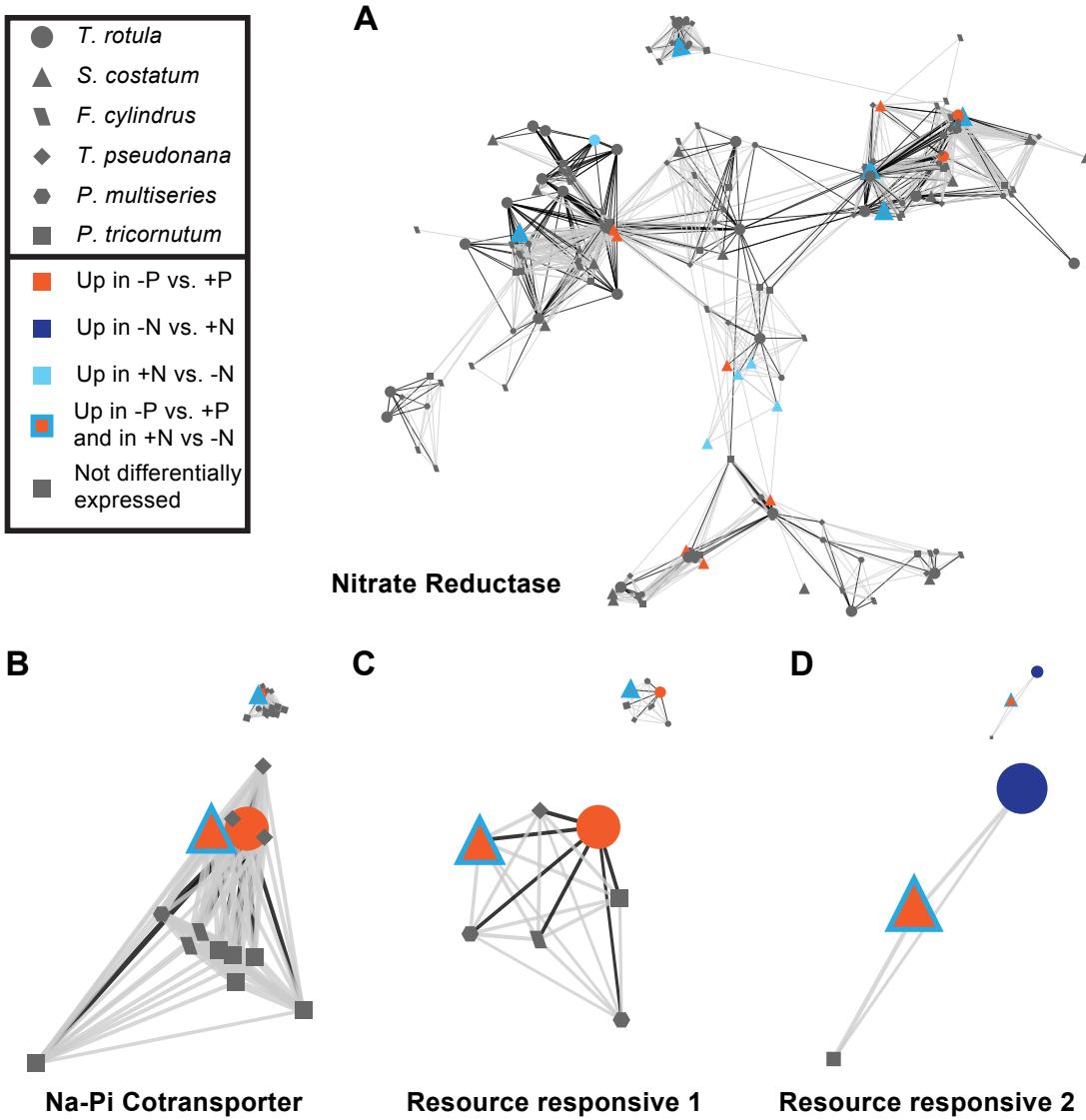


Figure A-9: Gene cluster known nutrient-responsive genes in *T. pseudonanana*: (A) assimilatory nitrate reductase and (B) sodium-phosphate cotransporter and novel resource-responsive (RR) gene families: (C) RR1 and (D) RR2. Transcripts from the transcriptomes of *T. rotula* and *Skeletonema* spp. were clustered based upon relative homology with available diatom genomes: *F. cylindrus*, *P. tricornutum*, *P. multiseries*, and *T. pseudonanana*. Symbols indicate different species, while color indicates regulation in the field incubation experiments. Two nodes within a gene cluster are connected by an edge if they share a homologous protein (reciprocal BLAST hit with a minimum of 1e-5 score and minimum 20% identity). Gene clusters are visualized using an edge-weighted spring-embedded model based on e-value, meaning that genes that are closer together are more similar. The width of the line correlates to the magnitude of the e-value, with lower e-values represented by thicker lines and higher e-values represented by thinner lines.

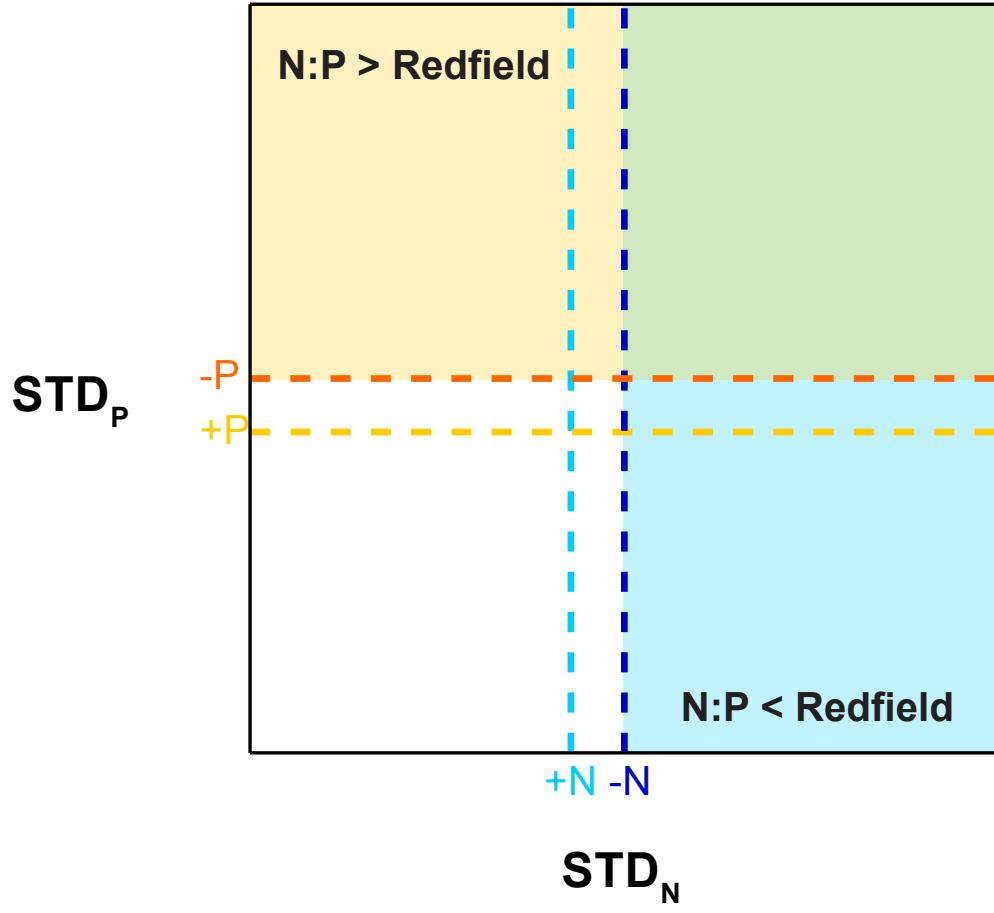


Figure A-10: A conceptual schematic of STD_N plotted against STD_P hypothesized regions of $N:P >$ Redfield physiology and $N:P <$ Redfield physiology highlighted.

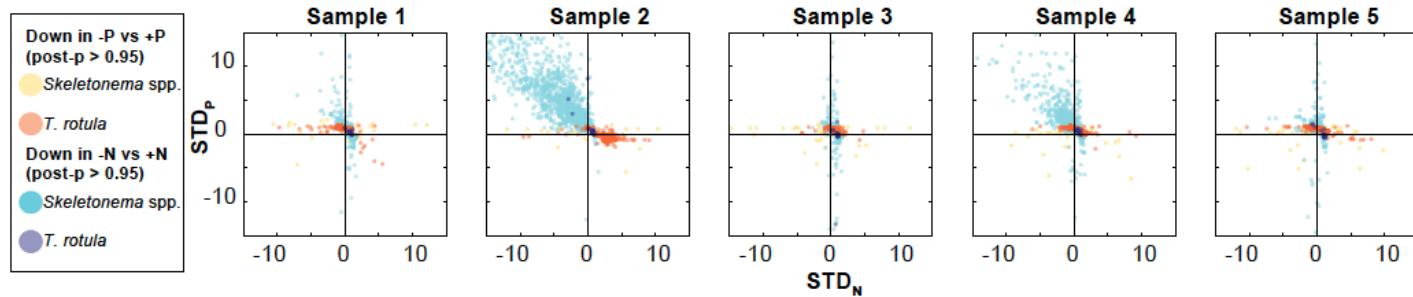


Figure A-11: Evolution of niche space indexing over time in Narragansett Bay for *T. rotula* and *Skeletonema* spp.. The stable gene normalized field signal from genes identified as significantly (2-fold change, post- $p > 0.95$) down-regulated in -P vs +P for *Skeletonema* spp. (yellow) and *T. rotula* (orange) and in -N vs +N for *Skeletonema* spp. (cyan) and *T. rotula* (dark blue) was proportionalized relative to the expression for those genes in nutrient incubations, yielding the STD_N and STD_P . These data are plotted for Sample 1 through Sample 5.

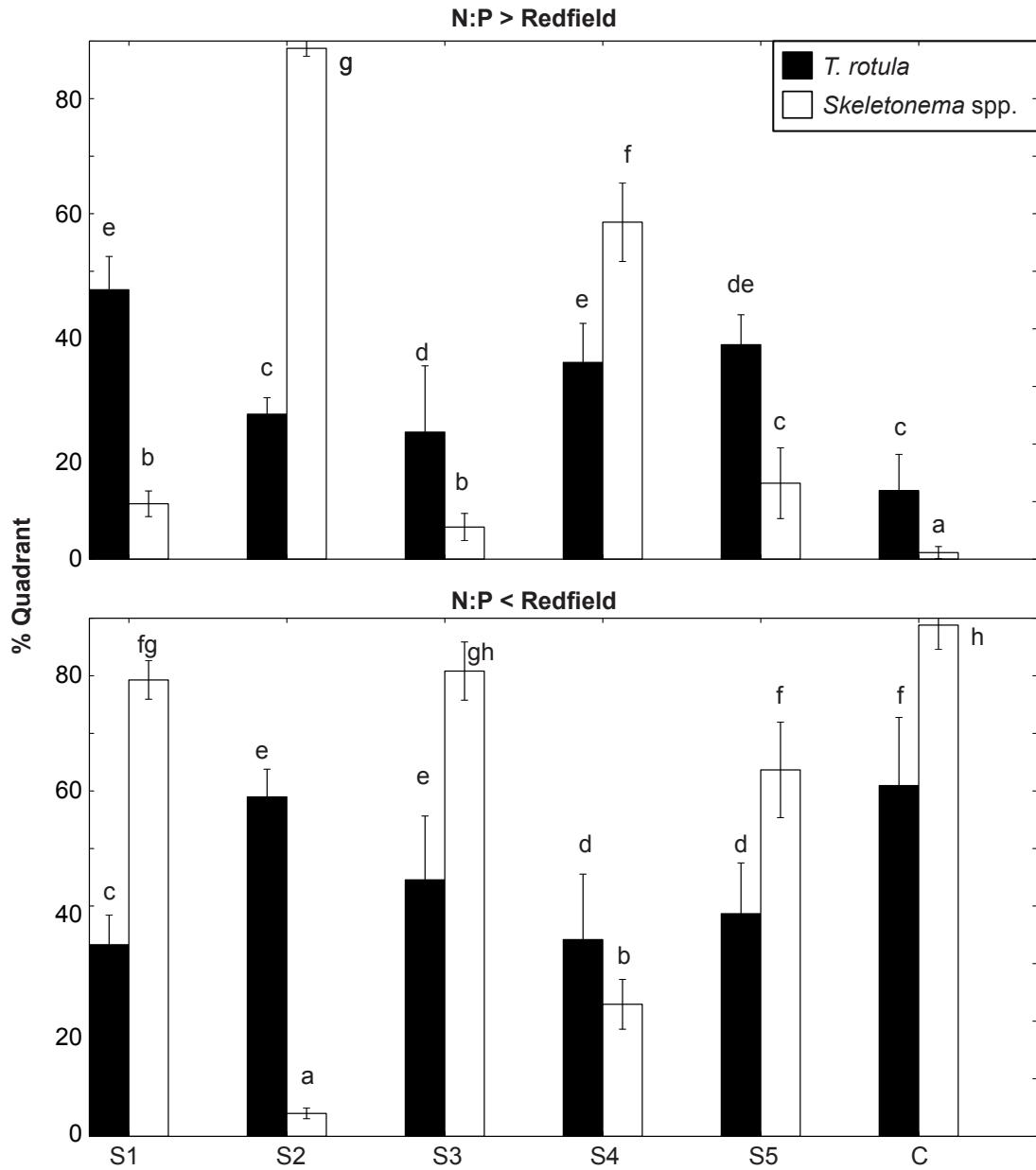


Figure A-12: The percentage of identified nutrient responsive genes falling into the N:P > Redfield and N:P < Redfield quadrants for *T. rotula* and *Skeletonema* spp.. The total number of genes falling into the N:P > Redfield quadrant ($STD_P > C$; $STD_N < C$, for $0.25 < C < 0.75$) and the N:P < Redfield quadrant ($STD_P < C$; $STD_N > C$, for $0.25 < C < 0.75$). The value of C was varied over 10 different values and the average percentages of genes falling into each of the quadrants is depicted above based on the size of the circle at the median STD_N and STD_P for the genes in the quadrant. Similarity of data between species by quadrant was assessed using an analysis of variance (ANOVA) with a generalized linear model. The results from a post hoc Tukey test show the divergence of species across time ($p < 0.05$).

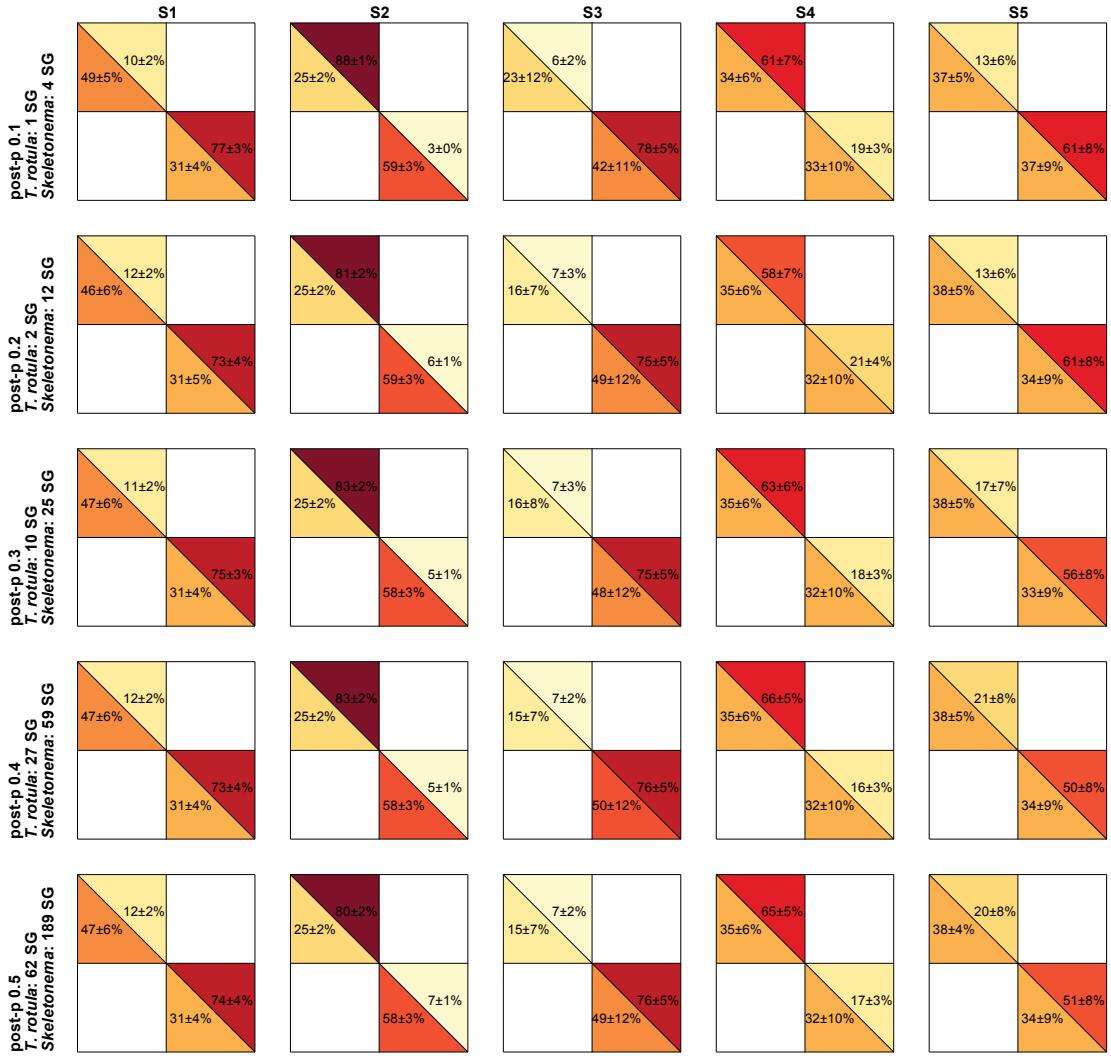


Figure A-13: The impact of stable gene selection on the quadrant localization of the resource responsive gene sets. The posterior probability cutoff used in the selection of stable genes was varied from 0.1 to 0.5 for a fold change of 1.25. The percentage of identified nutrient responsive genes falling into the N:P > Redfield and N:P < Redfield quadrants for *T. rotula* and *Skeletonema* spp. across the five sample points and five posterior probability values is depicted.

A.2.2 Supplemental Tables

Table A.1: My caption

Sample	Total library size (paired end reads)	Mapped representation in library		
		<i>T. pseudonana</i>	<i>T. rotula</i>	<i>S. costatum</i>
S1	89455034	2.98%	17.50%	33.50%
S2	64888267	0.41%	11.70%	54.90%
S3	103250243	0.39%	7.30%	9.00%
S4	45370867	0.68%	8.80%	8.30%
S5	55061692	0.88%	10.40%	11.20%
Ambient Control	51508197	0.27%	13.40%	8.00%
+N	58626239	0.43%	6.10%	5.30%
-N	44561851	0.41%	8.70%	8.30%
+P	51130364	0.29%	8.50%	8.00%
-P	58834022	0.40%	6.60%	6.50%

A.3 Appendix for Chapter 4

A.3.1 Supplemental Figures

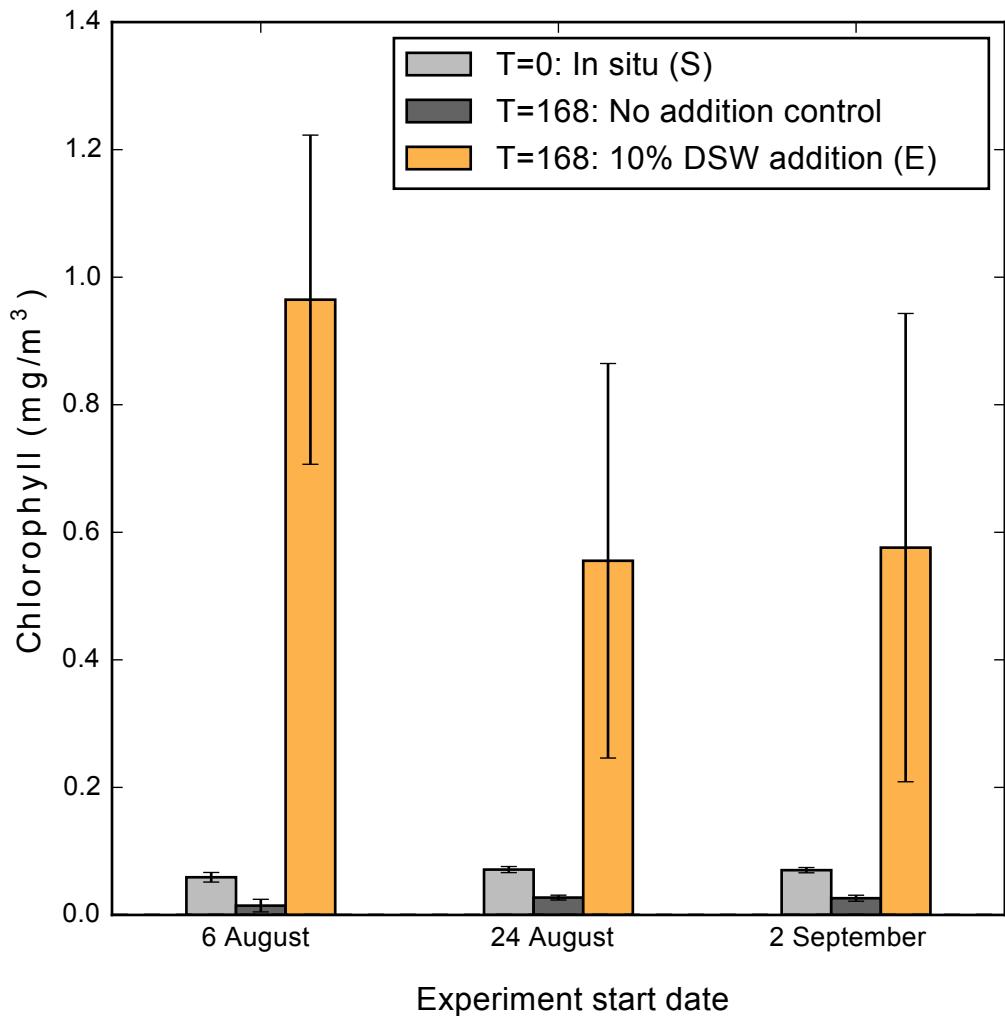


Figure A-14: Chlorophyll a of replicated experiments for *in situ* samples (S), a no addition control, and a 10% deep seawater (DSW) amendment (E). Incubation samples were harvested after 168 hours.

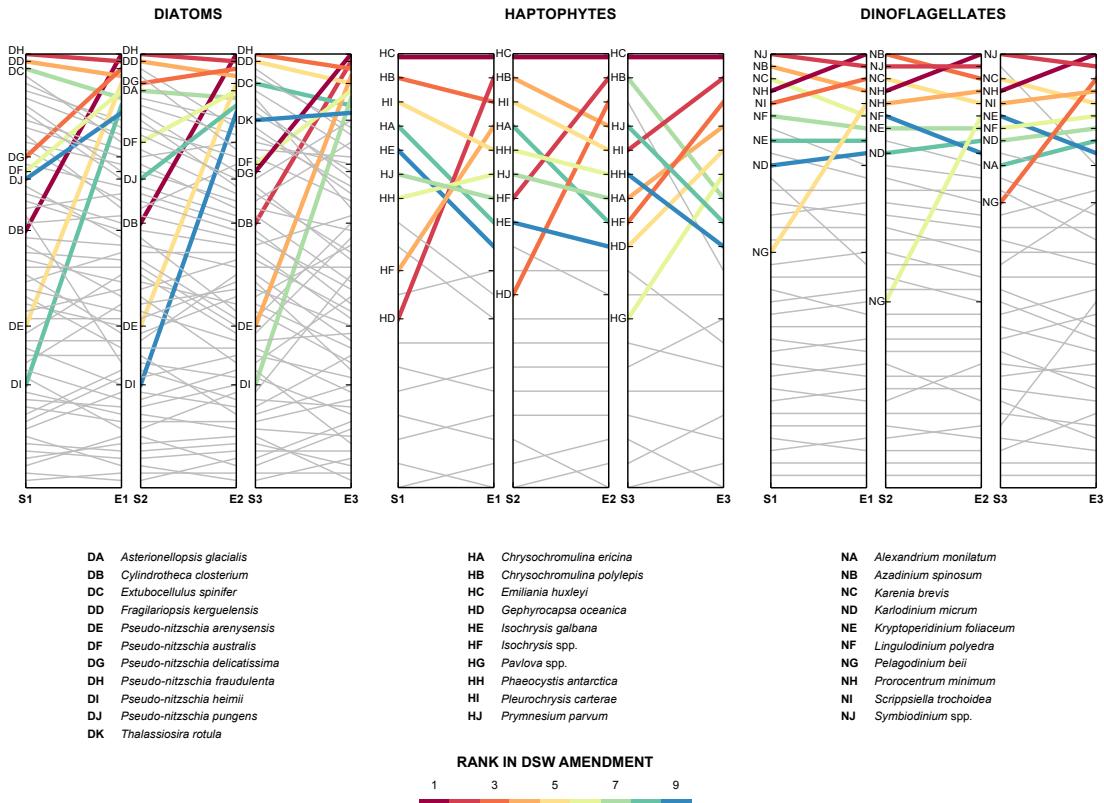


Figure A-15: Rank abundance shifts in the species composition of diatoms, haptophytes and dinoflagellates for the three experiments. The relative shift in rank abundance for each species is depicted for each incubation experiment (E1-E3) following deep seawater (DSW) addition. The nine most abundant taxa following DSW addition are highlighted for each of the functional groups. Although the species that recruited the reads are denoted here this is highly driven by the composition of the database and does not necessarily indicate the actual species present, but rather the closest species present in the database.

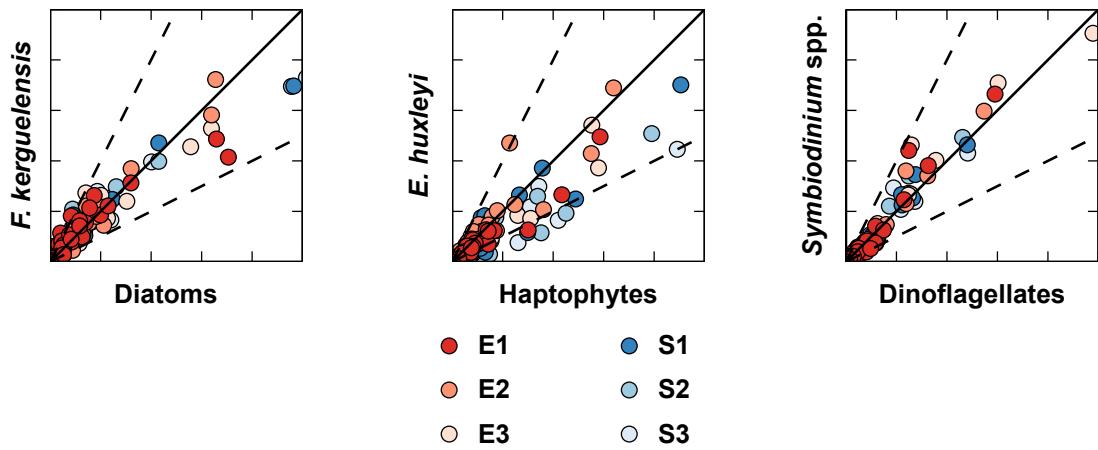


Figure A-16: Comparison of the quantitative metabolic fingerprint (QMF) between the whole functional group and representative taxa. The proportion of reads falling into each of the modules depicted in Figure 2 is plotted for S1-S3 and E1-E3, comparing the summed functional group signal and that of a representative taxon. Color of the marker indicates the sample; solid and dashed lines mark the 1:1 and 1:2 lines, respectively.

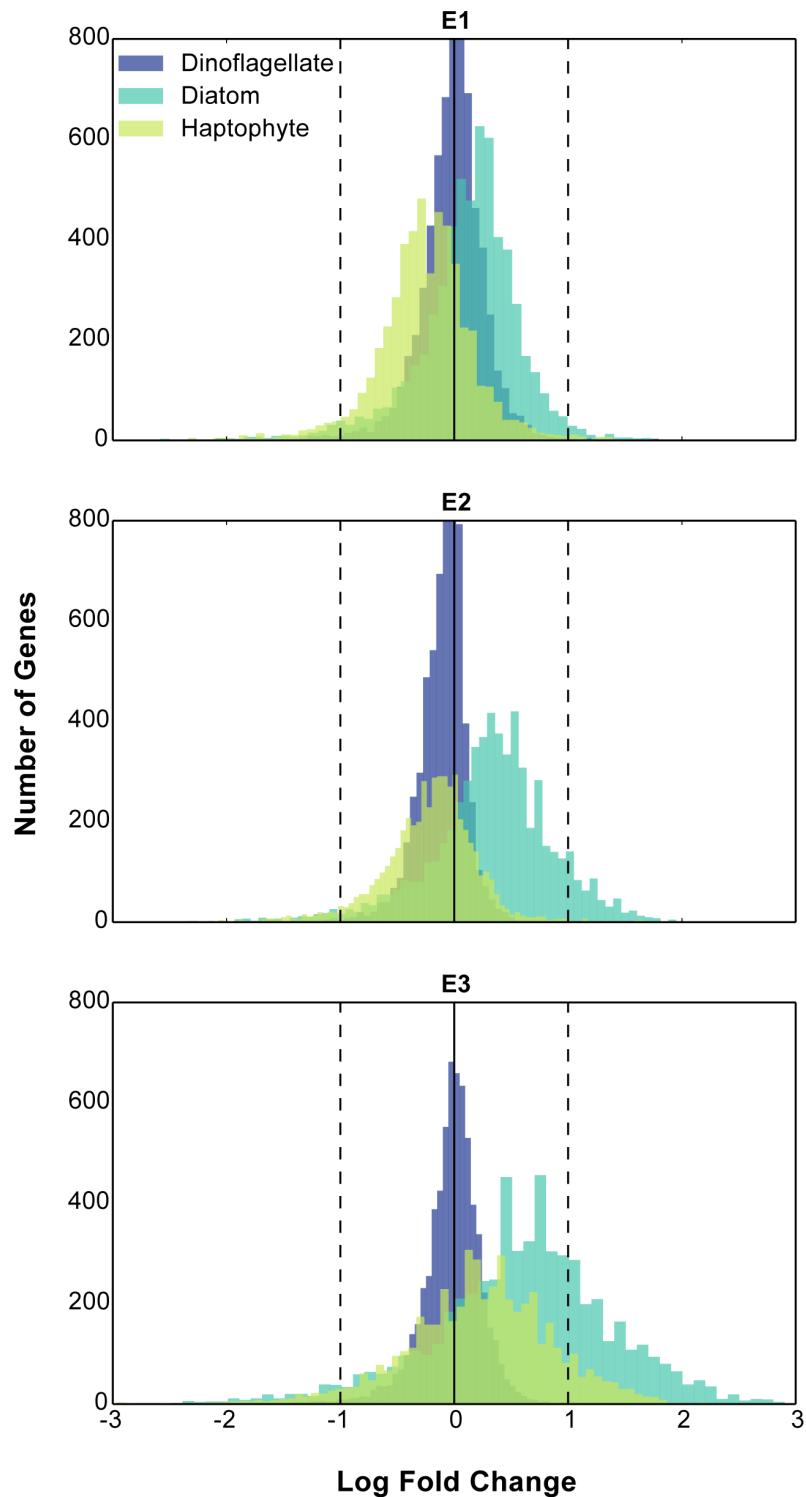
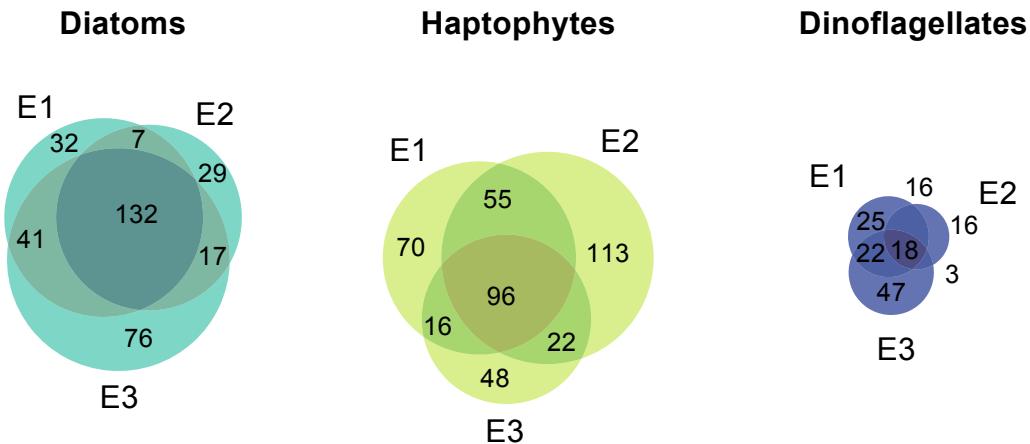


Figure A-17: Distribution of log fold change following deep seawater (DSW) addition. Histogram of the number of genes falling within each of the log fold change bins for diatoms, haptophytes and dinoflagellates. Solid line indicates no fold change; dashed lines indicate 2 fold-change both up and down.

Genes with Increased Transcript Abundance



Genes with Decreased Transcript Abundance

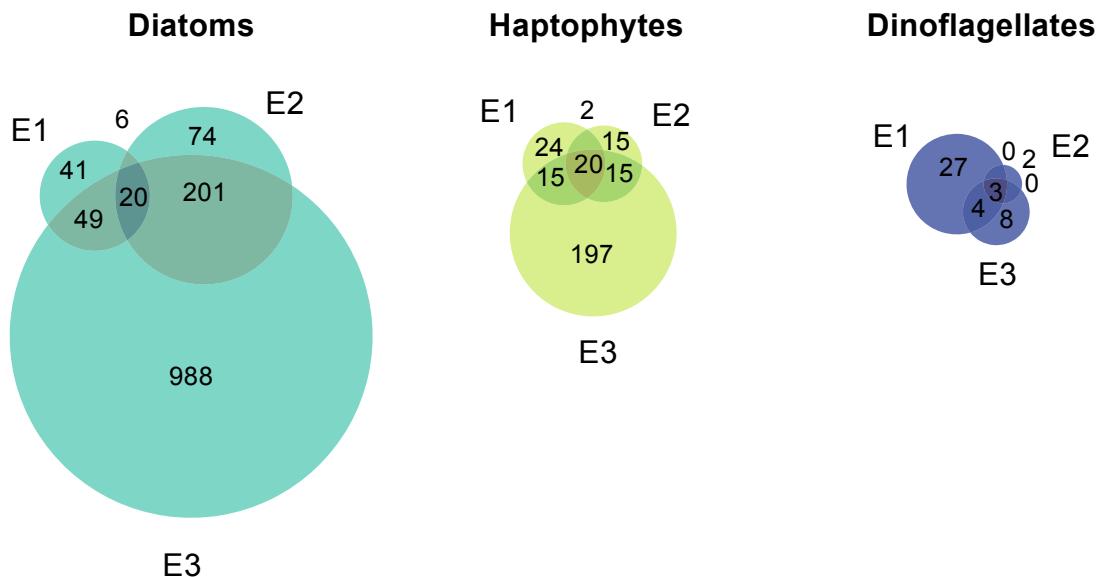


Figure A-18: Weighted Venn diagrams of genes with significantly different abundances following deep seawater (DSW) addition by functional group. The uniqueness of KEGG orthologs with increased or decreased abundances as determined by ASC (2 fold-change, post-p > 0.95) across experiments was assessed for diatoms, haptophytes, and dinoflagellates.

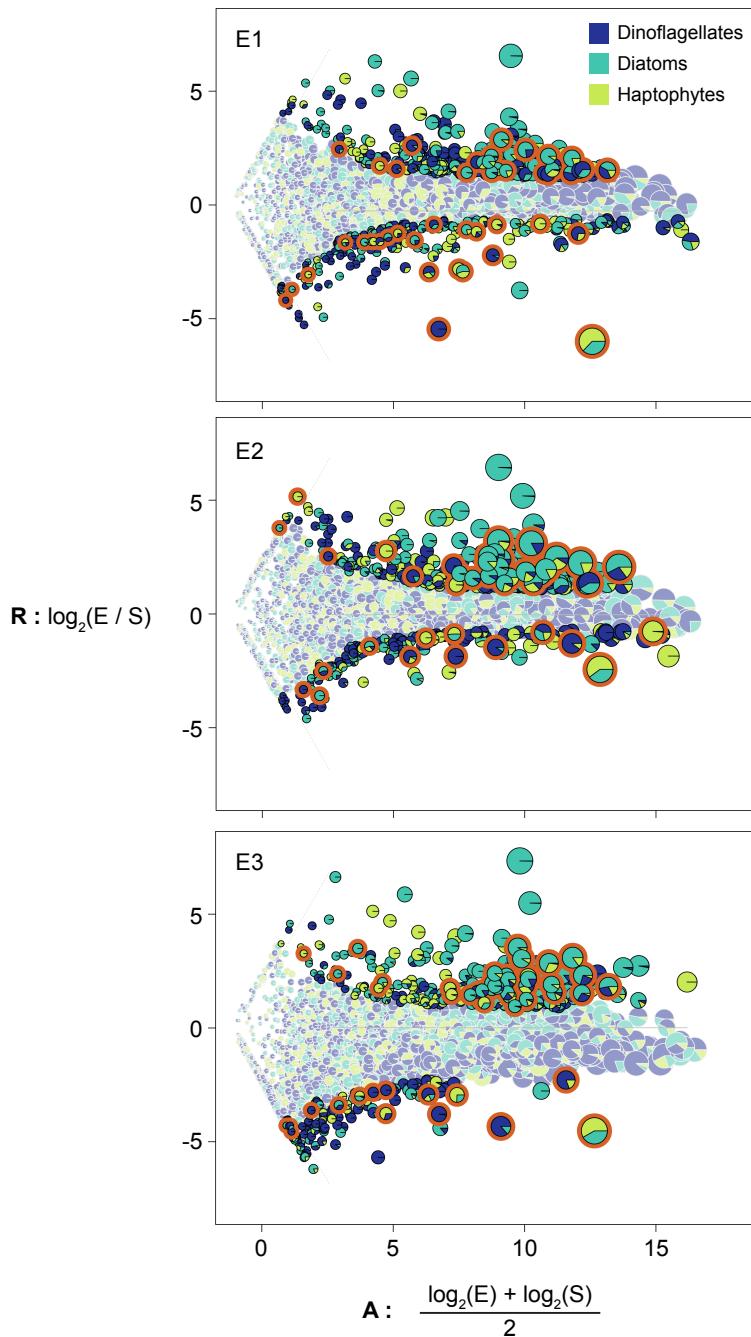


Figure A-19: Microbial Assemblage Normalized Transcript Analysis (MANTA) ratio-averaged plots for global shifts in expression of KEGG orthologs. Fold change ratio (R) and average read count (A) are plotted for read counts in the *in situ* (S) and deep seawater (DSW) amendment (E) samples across the three sample pairs (S1:E1, S2:E2, S3:E3). The trimmed mean of fold-change values is noted as a gray solid line; orthologs unique to one library are separated by gray dashed lines. Pies indicate the taxonomic distribution of orthologous reads across the three functional groups. KEGG orthologs that were significantly differentially expressed (DE) (adjusted $P > 0.05$) are outlined in black and those not significantly DE are outlined in gray. DE KEGG orthologs that fall in the Energy Metabolism KEGG module are outlined in orange.

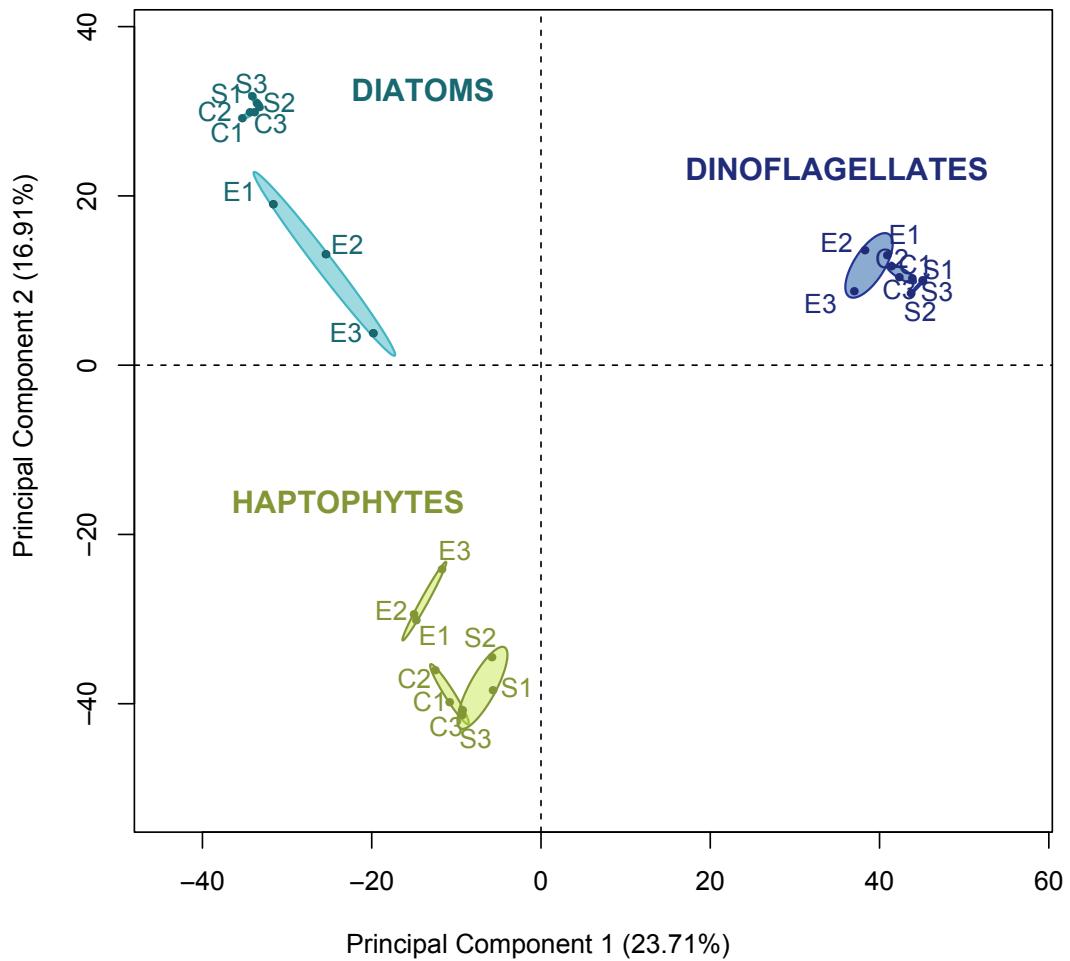


Figure A-20: Principal component analysis of the quantitative metabolic fingerprint (QMF) signals across *in situ*, no addition control, and deep seawater (DSW) amended samples. Principal component analysis of the QMF signals for each of the functional groups across *in situ* (S1-S3), control no addition (C1-C3) and DSW amendment (E1-E3); 95% confidence ellipses are indicated for each of the sample types by functional group.