Supplemental Information for :

Molecular underpinnings and biogeochemical consequences of enhanced diatom growth in a warming Southern Ocean

Loay Jabre[a], Andrew E. Allen[b,c*], J. Scott P. McCain[a], John P. McCrow[b], Nancy Tenenbaum[d], Jenna L. Spackeen[e], Rachel E. Sipler[e,f], Beverley R. Green[g], Deborah A. Bronk[e,h], David A. Hutchins[d*], Erin M. Bertrand[a,*#]

[a] Dept. of Biology, Dalhousie University, 1355 Oxford Street, PO BOX 15000, Life Sciences Center, Halifax, NS, Canada, B3H 4R2
[b] Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA
[c] Integrative Oceanography Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92037, USA
[d] University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA
[e] Virginia Institute of Marine Science, College of William & Mary, Gloucester Point, VA, 23062, USA
[f] Memorial University of Newfoundland, 0 Marine Drive, Ocean Sciences Centre, St. John's, NL, Canada, A1C 5S7
[g] Dept. of Botany, University of British Columbia, #3200-6270 University Boulevard, Vancouver, BC, Canada, V6T 1Z4
[h] Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME 04544, USA
[*] **Corresponding Authors.**
[#] **Corresponding Author for Manuscript submission.**

Corresponding Authors:
Erin M. Bertrand - erin.bertrand@dal.ca
Andrew E. Allen - aallen@jcvi.org
David A. Hutchins - dahutch@usc.edu

This document includes:

- Suplemental text
- Figures S1 – S10
- Tables S1 – S9
- Supplemental References

Full Materials and Methods:

Experimental Design – The planktonic microbial community was sampled at 3-m depth via diaphragm pump from the sea ice edge in McMurdo Sound, Antarctica on January 15th, 2015 (165°24.7985'E, 77°37.1370'S) from 11:00 - 12:15 using trace metal clean techniques previously described (1). In-situ water temperature at the time of sampling was -1 °C. The community was protected from light upon sampling using dark trash bags, stored at 0 °C until 17:00 and then split into trace-metal cleaned polycarbonate bottles (two 1.1L and one 2.7L per treatment), with and without iron supplementation at three different temperatures. Bottles were kept at -0.5 ± 0.2 °C, 3 ± 0.5 °C or 6 ± 0.5 °C at constant 65-85 uE m^{-2} sec^{-1} irradiance in indoor incubators for a total of 7 days. For iron supplmentation, 2 nM iron was added as $Fe(NO_3)_3$ from an ultrapure analytical standard solution, 1001 mg L^{-1} in 2% nitric acid. This was diluted to a working stock in pH 2.5 milli Q water with hydrochloric acid, resulting in a negligible nitrate addition to iron amended bottles.

Metatranscriptome Sampling and Assembly – Four separate metatranscriptome samples were taken from the initial community, one at the sea ice edge (IE; approximately 2L) and triplicates in the laboratory during bottle incubation setup (T0, approximately 2L). Subsamples of single replicates from each experimental treatment were harvested on January 16th (T1) and subsamples of each replicate (n = 3 for each experimental treatment) were taken again on January 20th (T5). Each was harvested onto 0.2 µm Sterivex^{TM} filters. RNA was extracted and sequenced via paired end Illumina HiSeq. Total RNA was extracted using Trizol reagent (Thermo Fisher Scientific). Ribosomal RNA was removed with Ribo-Zero Magnetic kits (Illumina). A mixed Removal Solution was prepared from plant, bacterial, and human/mouse/rat Removal Solution at a ratio of 2:1:1. The resulting rRNA subtracted RNA was purified and subjected to amplification and cDNA synthesis, using the Ovation RNA-Seq System V2 (TECAN). One microgram of the resulting high-quality cDNA pool was fragmented to a mean length of 200 bp, and libraries were prepared using Truseq kit (Illumina) from the -repair step in the manufacturers protocol and subjected to paired-end sequencing via Illumina HiSeq.

Illumina paired reads were filtered to eliminate primer sequences and quality trimmed to Phred Q33, and rRNA identified and removed using riboPicker (2) (average 13.5 % rRNA). Transcript contigs were assembled de novo, in a combined assembly, using CLC Assembly Cell (http://www.clcbio.com) and ORFs predicted using FragGeneScan (3). Reads were mapped to ORFs using CLC (73% read mapping), and ORFs were annotated for putative function using hidden Markov models and BLAST-p against PhyloDB (1). ORFs were filtered to eliminate those with low mapping coverage (< 50 reads total over all samples), proteins with no BLAST hits, and no known domains (See Tables S7, S8, S9). The remaining set of ORFs were assigned to chloroplast, mitochondrial or nuclear origin based on the best BLAST-p hit above e-value $1e^{-3}$ to an organism with known organellular peptide sequences (nuclear by default), and used for further comparative analysis.

Taxonomic groups of interest were defined (Fig. S2) and each ORF was assigned to a group based on best LPI taxonomy (1, 4). A total of 64,487 ORFs were assigned to *Fragilariopsis* and 28,650 ORFs were assigned to *Pseudo-nitzschia*. Reads per kilobase mapped (RPKM) expression values for each ORF were calculated and taxon normalized using a normalization factor representing the summed taxonomic group contribution to total nuclear-assigned reads per sample. For example, the taxon-normalized expression of an ORF assigned to *Fragilariopsis* in a particular library is given as reads mapped to that ORF/ORF length/total *Fragilariopsis* nuclear assigned reads in that library. ORFs were clustered into orthologs and protein families using MCL (5). MCL clustering was run in label mode (parameter -abc), with the default inflation setting 2.0 (parameter -I), on ratios of best BLAST-p bitscore to self-hit bitscore using BLASTALL (e-value: $1e^{-3}$). Group normalized cluster (average/total) RPKM expression values were calculated by pooling the taxon-normalized expression values for each group within a cluster. These taxon-normalized RPKM values were used to examine gene and cluster abundance. Cluster annotations were aggregated by annotation type (Kegg, KO, KOG, KOG class, Pfam, TIGRfam, EC, GO) and a single

annotation chosen to represent each cluster based on the lowest Fisher's exact test p-value (fisher.test in R) given the 2-way contingency table for each annotation coverage of each cluster.

***Pseudo-nitzschia* BLAST-p analyses:** A subset of TFG peptide sequences, assembled from meta-transcriptome sequences and annotated as most-likely from *Pseudo-nitzschia* (n= 27,420), as described above, were used in BLAST-p analysis against peptide databases from culture-based transcriptomes of several *Pseudo-nitzschia* species (Table S3). All the peptide matches that passed an E value threshold of 1E-05 were collected separately for each *Pseudo-nitzschia* species. Their distribution profiles in percent identity and the alignment length are summarized in Table S3.

18S rRNA Sequencing – 1-10 ng of total RNA was used to generate cDNA using the Life Technologies SuperScript III First Strand Synthesis system with random hexamer primers. The cDNA was diluted 10-fold and had final concentrations ranging from 10-50 ng/ul. Amplicon libraries targeting the V9 region of the 18S gene were generated as described here: https://www.protocols.io/view/amplicon-library-preparation-bmuck6sw. Briefly, cDNA was amplified via a one-step PCR using the TruFi DNA Polymerase PCR kit (Azura, Raynham, MA, USA). The 1389F (TTGTACACACCGCCC) and 1510R (CCTTCYGCAGGTTCACCTAC) primer set was used (6). Each reaction was performed with an initial denaturing step at 95°C for 1 minute followed by 30 cycles of 95°C for 15 seconds, 56°C for 15 seconds, and 72°C for 30 seconds. 2.5 µL of each PCR reaction was ran on a 1.8% agarose gel confirm amplification. PCR products were purified using Beckman Coulter AMPure XP beads following the standard 1x PCR clean-up protocol. PCR quantification was performed in duplicate using Invitrogen Quant-iT PicoGreen dsDNA Assay kit. Samples were then pooled in equal proportions followed by another 0.8x AMPure XP bead purification. The Pool was evaluated on an Agilent 2200 TapeStation and quantified with Qubit HS dsDNA. Sequencing was performed at the University of California, San Diego Sequencing Core on a single Illumina MiSeq lane (2 x 150bp) with a 15% PhiX spike-in.

Amplicons were analyzed with QIIME2 v2019.4 (7). Briefly, demultiplexed paired-end reads were trimmed to remove adapter and primer sequences with cutadapt (8). Trimmed reads were then denoised with DADA2 to produce amplicon sequence variants (ASVs) (9). Each pool was denoised with DADA2 individually to account for different error profiles in each run. Taxonomic annotation of ASVs was conducted with the q2-feature-classifier classify-sklearn naïve-bayes classifier (10, 11) PR^2 v4.12.0 (12) for 18S amplicons.

ISIP1 taxonomic re-assignment – *Pseudo-nitzschia* and *Fragilariopsis* ISIP1 genes could not be differentiated using BLAST-p. Instead, nucleotide ISIP1 sequences were compared to a reference database of ISIP1 genes from *F. cylindrus*, *F. kergulensis*, *P. granii*, *P. heimii*, *P. multistriata* and *P. fraudulenta* using BLAST-n. ISIP1 sequences were assigned to *Fragilariopsis* or *Pseudo-nitzschia* based on lowest e-value score. These sequences were then manually placed in clusters and their abundance was normalized to each taxon as previously described.

Query for Domoic Acid Biosynthesis (DAB) Genes – BioEdit v7.2 was used to conduct a local BLAST-p search for DAB genes in our data using Blosum62 similarity matrix. Amino acid query sequences from *Pseudo-nitzschia multiseries* DAB-A (GenBank: AYD91073.1), DAB-B (GenBank: AYD91072.1), DAB-C (GenBank: AYD91075.1) and DAB-D (GenBank: AYD91074.1) (13) were used.

LHC assignments – All *Fragilariopsis* and *Pseudo-nitzschia* ORFs annotated as 'chlorophyll binding' or 'light harvesting proteins' were selected for further inspection. For *Fragilariopsis* LHCs, a BLAST-p was performed against *F. cylindrus* CCMP1102 and annotations were retrieved for the top BLAST hit for each amino acid sequence. *Pseudo-nitzschia* LHC sequences were identified by comparison with those collected during the annotation of the *Pseudo-nitzschia multiseries* CLN-47 genome. The protein sequences of both diatoms were assigned to the Lhcf, Lhcr, Lhcx or Lhcz groups following a previous diatom LHC classification based on maximum-likelihood phylogenetic trees and published as Supp. Information 11 and Supp. Fig. 20 of Mock et al. (2017) and in Hippmann et al. (2017) (14, 15). The dominant Lhcf clade was further

subdivided into Lhcf_I, Lhcf_II (diatom-specific), and Lhcf_III groups. PID numbers and the clusters to which they were assigned (see above) can be found in Table S6.

Plastocyanin Tree – Previously identified plastocyanin sequences were retrieved from MMETSP (*Fragilariopsis kerguelensis*_0735, *Pseudo-nitzschia heimii*_1423, *Coscinodiscus wailesii*_1066), JGI (*Fragilariopsis cylindrus*_272258), NCBI (*Thalassiosira oceanica*_EJK71623.1), Moreno et al. 2018 (*Pseudo-nitzschia subcurvata*) (16), and Cohen et al. 2018 (*Pseudo-nitzschia granii*) (17), and aligned with plastocyanin sequences from *Pseudo-nitzschia* and *Fragilariopsis* in our dataset using Clustal Omega in SeaView v5.0 (Fig. S8). A maximum-likelihood phylogeny was then estimated using PhyML with LG model and 100 bootstrap iterations in SeaView v5.0, and the resulting tree was edited using FigTree v1.4.4.

Nutrient Measurements – Samples from initial (T0) and subsequent time points (T1, T3, T5, and T7) were collected, passed through a GF/F filter (Whatman; 0.7 µm nominal pore size; combusted at 450 °C for 2 hours), and filtrate was stored frozen (-40 °C) until further analysis. A Lachat QuickChem 8500 autoanalyzer was used to measure duplicate concentrations of dissolved nitrate, phosphate and silicate (detection limit 0.03 µmol N L^{-1}, 0.03 µmol P L^{-1} and 0.05 µmol Si L^{-1}; (18)). Samples for ammonium, collected on T0 and T7, were measured in triplicate on a Shimadzu UV-1601 spectrophotometer using the manual phenol-hypochlorite method (detection limit 0.05 µmol N L^{-1}; (19)).

Uptake Measurements – Nitrate and bicarbonate uptakes were assessed using the initial community (T0) collected from the ice edge and during T1, T3, and T7 of the experiment (Fig. 1 and Fig. S1). Uptake rates were measured using ^{15}N and ^{13}C stable isotope tracer techniques, and substrates used included ^{15}N-labeled potassium nitrate ($K^{15}NO_3^-$; 98%) and ^{13}C-labeled sodium bicarbonate ($NaH^{13}CO_3^-$; 99%; both substrates came from Cambridge Isotope Laboratories, Andover, MA). Uptake experiments at T0 were done in triplicate using 1 L polyethylene terephthalate glycol-modified bottles. At T1, T3, and T7 a single replicate of each treatment was subsampled, and uptake experiments were done in duplicate in 230 mL polycarbonate conical

bottles (all bottles were acid washed with 10% hydrochloric acid and thoroughly rinsed with ultrapure water). After tracer level additions (less than 10% of background concentrations) of ^{15}N and ^{13}C-labled substrates were made, bottles were returned to their respective incubators for approximately 6 hours. Incubations were terminated by filtering microbial communities (> 0.7 µm) onto combusted (450 °C for 4 hours) Whatman GF/F filters. During T0 incubations, two microbial size fractions (0.7 – 5.0 µm collected on GF/F filters and > 5.0 µm collected on Sterlitech silver membrane filters) were added together to represent the > 0.7 µm microbial community. Filters were kept fozen (-40 °C) inside 1 mL cryo vials until particulate nitrogen and carbon concentrations and isotopic enrichment of ^{15}N and ^{13}C were measured on a Europa 20/20 isotope ratio mass spectrometer. Absolute uptake rates for ^{15}N-labeled nitrate and ^{13}C-labeled bicarbonate were calculated according to (20, 21) respectively. Nitrate uptake rates were not corrected for isotope dilution because concentrations of nitrate were greater than 5.5 µmol N L^{-1} at all time points, and isotope dilution is generally negligible when concentrations are high (22).

Cell Counts – Phytoplankton cell count samples from the initial (T0) and final days (T7) of the experiment were preserved with 1% glutaraldehyde, stored refrigerated in the dark and later enumerated in the lab on a Sedgwick Rafter counting chamber using an inverted compound light microscope (Accu-Scope 3032), as in (23). All plankton taxa were identified to the lowest taxonomic level possible according to (24, 25), with special attention to the diatom genera *Fragilariopsis* and *Pseudo-nitzschia*.

Statistical Analysis – We analyzed differential gene expression at T5 within observed taxa to determine which genes are responsive to Fe and temperature treatments in each group. First, we normalized reads mapped to each ORF by the abundance of nuclear reads assigned to that taxonomic group in total, which controls for changes in community composition across treatments. We then used a generalized linear model with one categorical explanatory variable, with each category representing a unique experimental treatment. To examine the effect of Fe, temperature, and their interaction on gene expression, we specified model contrasts. For Fe, we

tested the difference between the sum of coefficient estimates for all Fe treatments, minus the sum of coefficient estimates for all -Fe treatments. We followed a similar approach for temperature, where we tested the difference between the sum of coefficient estimates for one temperature treatment versus another temperature treatment. For both approaches, we divided these differences by the number of treatments in the sum (i.e. 3 for Fe test and 2 for each temperature test). To test for statistical significance, we used empirical Bayes quasi-likelihood F-tests (glmQLFTest in edgeR). To examine the interaction between Fe and temperature, we set up a contrast to compare the difference between +/-Fe treatments at constant temperature, and then compared this difference to a distinct temperature treatment. The test for a significant interactive effect is based on the difference of these differences – i.e. is the expression of a gene to Fe altered due to warming? Throughout, we used a p-value cut off of 0.05 for statistical significance.

To examine the effects of iron on gene differential expression and fold-change magnitude, -Fe treatments for all temperatures at T5 (-Fe -0.5 °C, -Fe 3 °C, -Fe 6 °C) were compared to +Fe treatments for all temperatures at T5 (+Fe -0.5 °C, +Fe 3 °C, +Fe 6 °C). For temperature effect, -0.5 °C was compared to 3 °C at all iron conditions at T5 (+Fe and -Fe at -0.5 °C vs +Fe and -Fe at 3 °C) and -0.5 °C was compared to 6 °C at all iron conditions at T5 (+Fe and -Fe at -0.5 °C vs +Fe and -Fe at 6 °C). For a gene to be considered upregulated with temperature, it had to be upregulated in both -0.5 °C vs 3 °C and -0.5 °C vs 6 °C (same rule applied for down regulation). If a gene is upregulated in one temperature comparison and downregulated in the other, it was not considered differentially expressed. Fold-change for temperature effect was calculated from -0.5 °C (+Fe and -Fe) vs 6 °C (+Fe and -Fe).

K.O. term enrichment analysis was performed on significantly upregulated (enriched) and downregulated (depleted) ORFs separately using KEGG enrichment functions in the GOstats R package (26). A hypergeometric distribution test was used to test for significant enrichment and

193 depletion at $p < 0.05$. K.O. terms associated with less than 10 ORFs were excluded from the
194 results.

195 **Supplemental Results and Discussion:**

196 ***Cobalamin Metabolism –*** Exogenous vitamin B_{12} (cobalamin) acts as a cofactor in the cobalamin-
197 requiring methionine synthase enzyme (MetH) found in all diatoms to synthesize the essential
198 amino acid methionine and facilitate one-carbon metabolism (27). Under low cobalamin
199 availability, however, certain diatoms including *Fragilariopsis* but not *Pseudo-nitzschia,* are
200 capable of synthesizing methionine from homocysteine using a less efficient cobalamin-
201 independent methionine synthase (MetE) (27, 28). Diatoms also upregulate CBA1, a cobalamin
202 acquisition related protein, under cobalamin deprivation (27, 28). Following the initial 24-hour
203 incubation period, CBA1 and MetH were upregulated in *Pseudo-nitzschia* after warming at 6 °C
204 (Fig. S7), but were not significantly influenced by iron or temperature following 5-day incubations
205 (Fig. 3; S7). In contrast, MetE and CBA1 transcripts were significantly upregulated in *Fragilariopsis*
206 following iron additions after 5-day incubations (Fig. 3; S7), suggesting rearranged metabolism to
207 cope with cobalamin deprivation that likely emerged in response to iron addition (1).

208 Rapid cobalamin uptake by *Pseudo-nitzschia*, facilitated by upregulation of CBA1 after 24
209 hours, may give it a competitive advantage when cobalamin is available. However, the ability of
210 *Pseudo-nitzschia* to maintain vigorous growth without significantly elevating CBA1 expression
211 after 5 days is notable, and suggests that these two diatoms may employ different strategies to
212 cope with low cobalamin availability. *Fragilariopsis* appears to reduce cobalamin demand
213 through the use of MetE while increasing investment in acquisition with CBA1. In contrast,
214 *Pseudo-nitzschia* may 1) have a MetH enzyme that is more efficient at elevated temperatures
215 compared to *Fragilariopsis*, 2) employ salvage and repair of degraded cobalamin complexes (17)
216 or 3) rely on bacteria in close physical association for cobalamin supply. Our data show that the
217 overall expression of genes encoding cobalamin salvage and remodeling proteins were not
218 influenced by iron status or temperature in *Pseudo-nitzschia* (Fig. S7), despite previous evidence

of their upregulation with iron addition in North Pacific diatoms including *P. granii* (17, 29). Further work comparing cobalamin uptake and methionine synthase kinetics between *Pseudo-nitzschia* and *Fragilariopsis,* and examining their relationships with cobalamin producing and consuming bacteria, could provide further insight into how these taxa cope with episodic decreases in cobalamin availability.

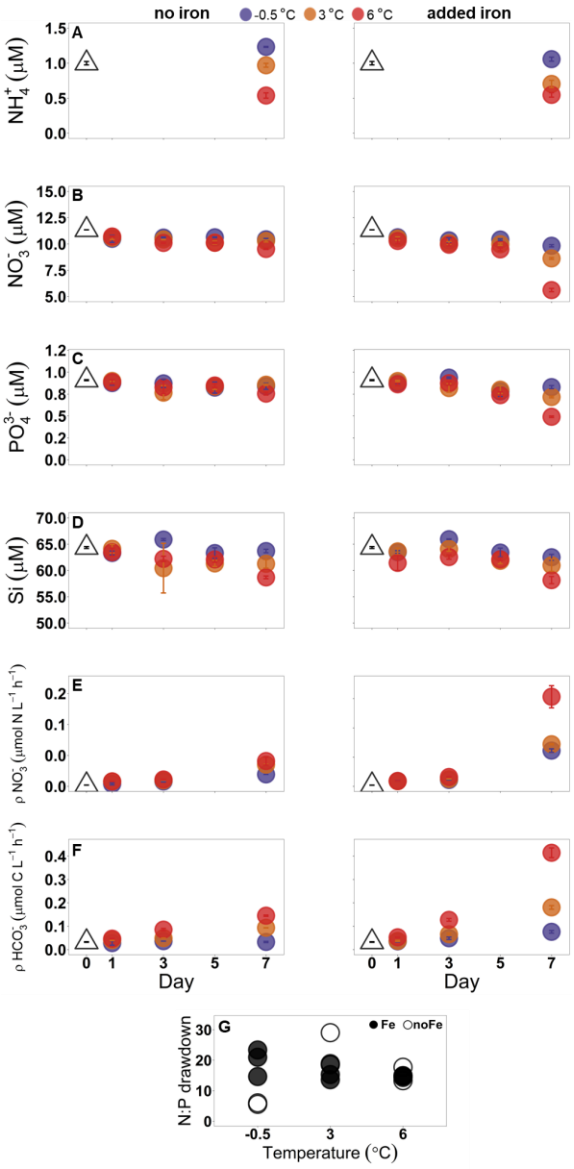

Figure S1 – A-D) Dissolved nutrient concentrations prior to incubations (day 0), and after 1, 3, 5 and 7 days of incubation with and without added iron at -0.5, 3, and 6 °C. Each point represents a mean value, where n = 2 on day 1, and n=3 on days 0, 3, 5, 7. In A-J, error bars represent ± 1 SD, and fall within the bounds of the symbol when not visible. **E,F)** Nitrate and bicarbonate uptake rates prior to incubations (day 0), and after 1, 3, and 7 days of incubation with and without added iron at -0.5, 3, and 6 °C. Each point represents a mean value, where n = 2 on days 1, 3, 7 and n = 3 on day 0. **G)** Dissolved nitrogen (nitrate + ammonium) : phosphate drawdown ratio at T7. Draw down is calculated as the difference in concentration between T7 and T0.

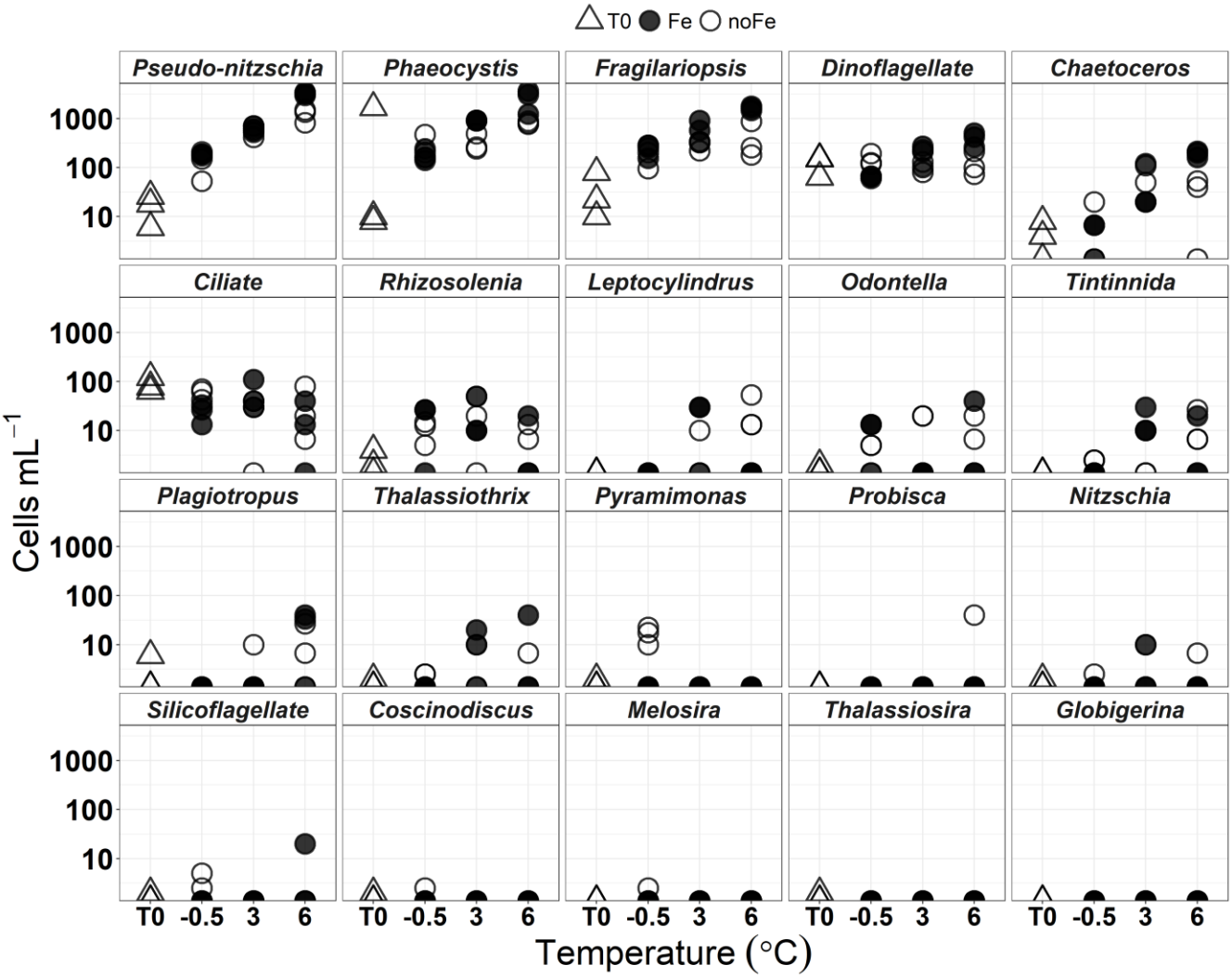

Figure S2 – Triplicate cell count measurements of initial (T0) samples and after 7-day incubations with and without iron addition at -0.5, 3 and 6 °C. Cells from the various taxonomic groups were counted and identified using light microscopy. The large variation in *Phaeocystis* cell counts at T0 could have resulted from difficulties in enumeration of colonial and single cell forms in the samples, and variability in colony abundance and size in the relatively small-volume samples used for microscopic cell counts.

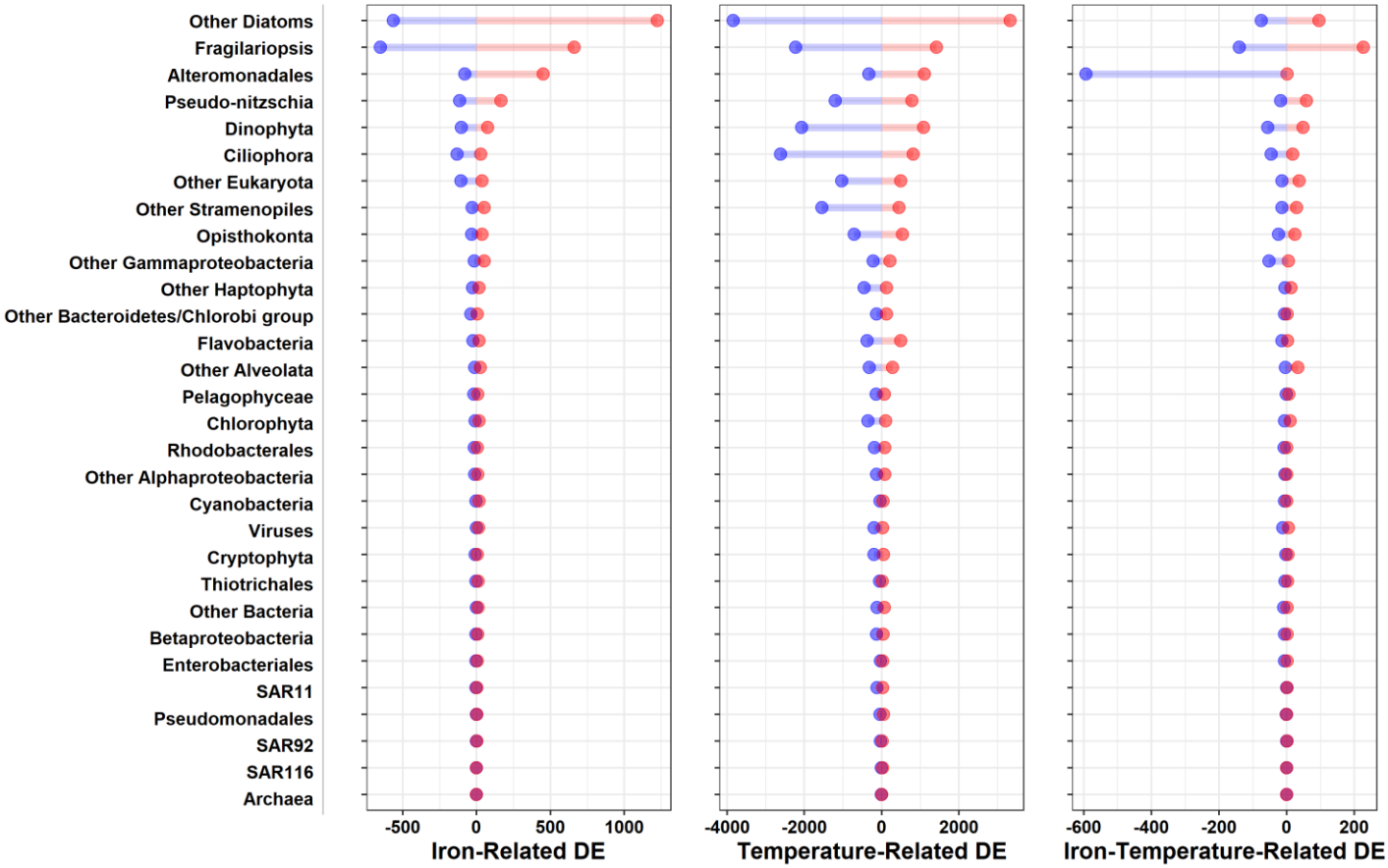

Figure S3 – Number of significantly differentially expressed open reading frames (ORFs) belonging to the 30 taxonomic groups identified in the metatranscriptome dataset. Red = upregulation, blue = downregulation. Differential taxon-normalized expression was calculated using the quasi likelihood test (glmQLFTest) in EdgeR, and p-value cut-off of 0.05 was used for statistical significance. Iron-related DE represents differential expression patterns observed with and without iron addition at T5, temperature-related DE represents differential expression patterns observed due to warming at T5, Iron-Temperature-related DE represents interactive iron-temperature effect on gene expression (Methods).

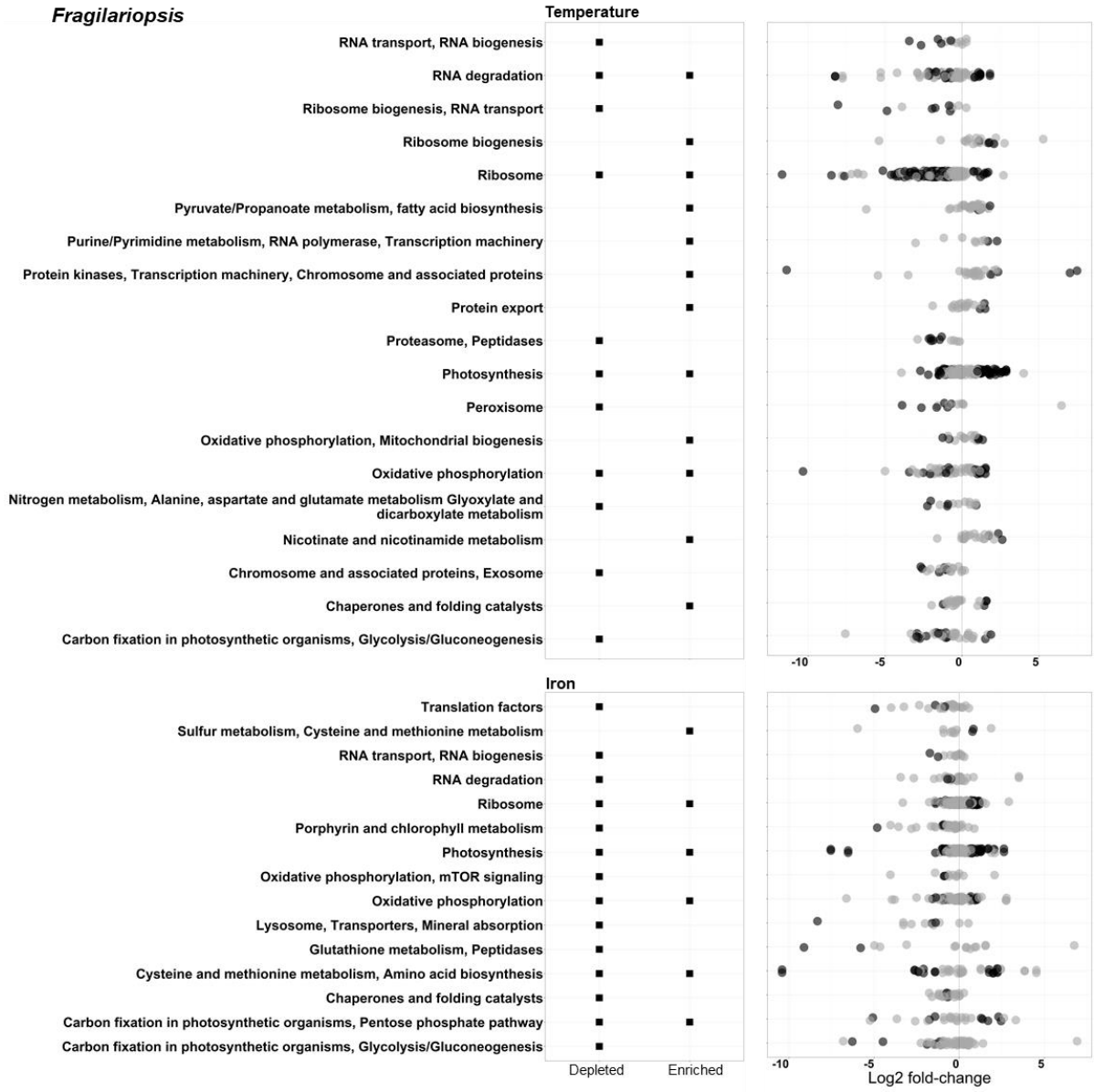

Figure S4 – KEGG Orthology (K.O.) term enrichment analysis using all *Fragilariopsis* ORFs that were annotated with a K.O. number. Black squares correspond to 'C' level annotations that were significantly ($p < 0.05$) upregulated (Enriched) and/or downregulated (Depleted) at T5 by temperature increase or iron addition. Circles correspond to the individual ORFs used in the analysis for each annotation. Black circles represent statistically significant ($p < 0.05$) up or down regulated ORFs (positive and negative Log_2 fold-change values, respectively). Temperature fold change was calculated using -0.5 °C vs 6 °C treatments. Iron fold-change was calculated using -Fe vs +Fe treatments at all temperatures (Methods).

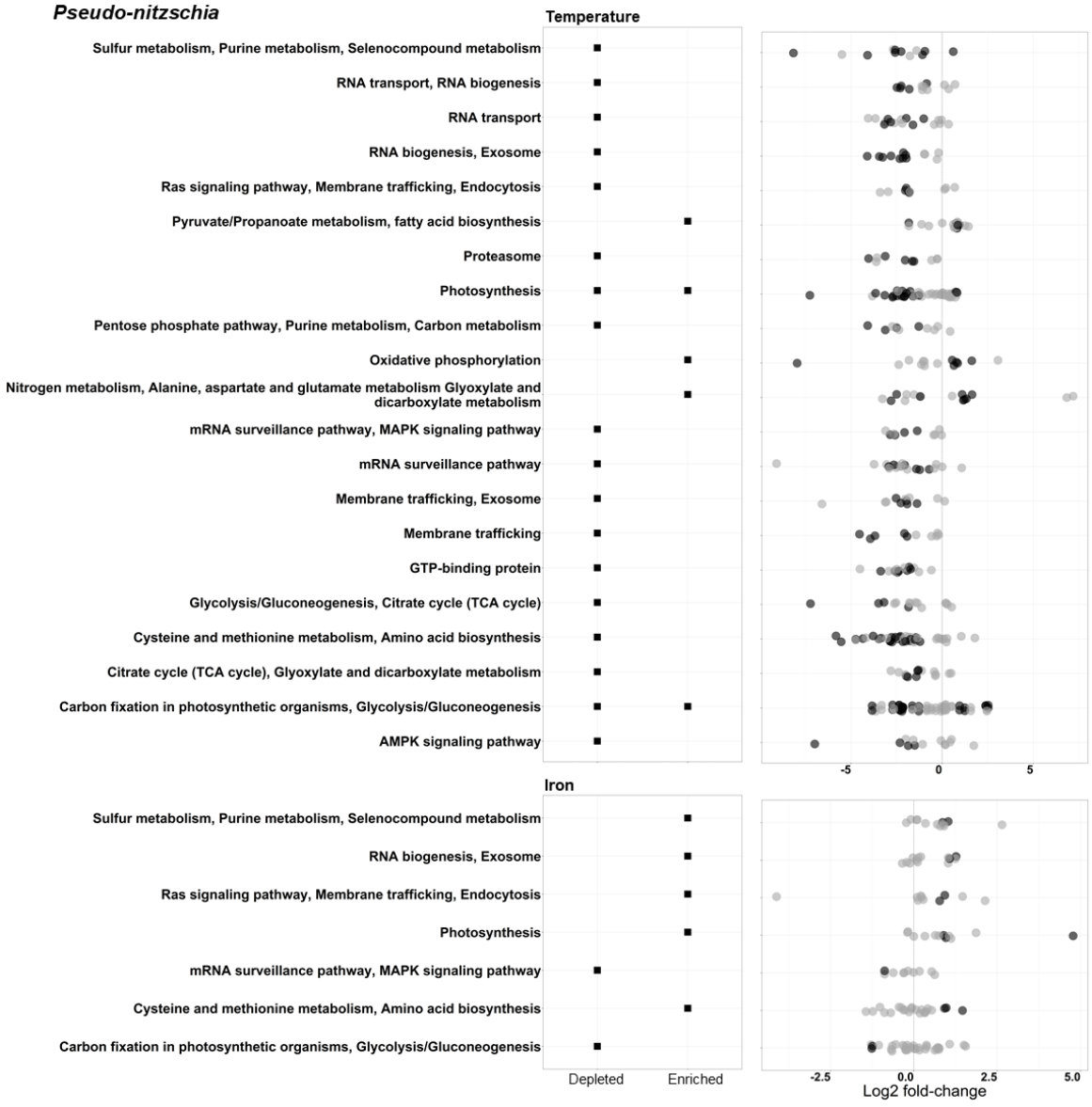

Figure S5 – KEGG Orthology (K.O.) term enrichment analysis using all *Pseudo-nitzschia* ORFs that were annotated with a K.O. number. Black squares correspond to 'C' level annotations that were significantly ($p < 0.05$) upregulated (Enriched) and/or downregulated (Depleted) at T5 by temperature increase or iron addition. Circles correspond to the individual ORFs used in the analysis for each annotation. Black circles represent statistically significant ($p < 0.05$) up or down regulated ORFs (positive and negative Log_2 fold-change values, respectively). Temperature fold-change was calculated using -0.5 °C vs 6 °C treatments. Iron fold-change was calculated using -Fe vs +Fe treatments at all temperatures (Methods).

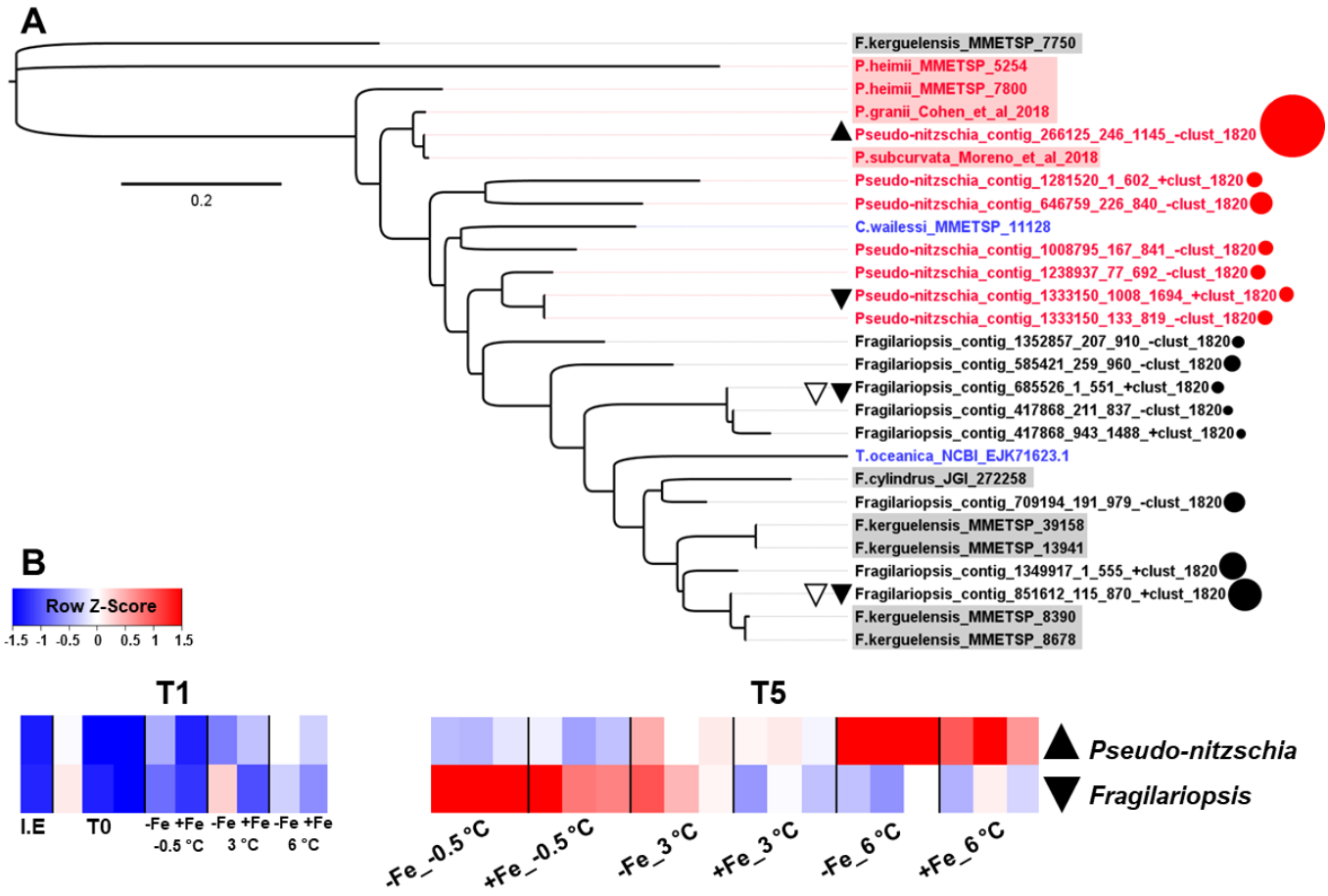

Figure S6 – A) Phylogenetic tree of the plastocyanin sequences (ORFs) comprising the plastocyanin MCL cluster from both *Pseudo-nitzschia* (red) *and Fragilariopsis* (black). Non-highlighted branch tips represent ORFs in our dataset, with corresponding point size representing mean taxon-normalized ORF expression. Highlighted branch tip labels represent previously identified plastocyanin sequences retrieved from the MMETSP dataset (*Fragilariopsis kerguelensis*_0735, *Pseudo-nitzschia heimii*_1423, *Coscinodiscus wailesii*_1066), JGI (*Fragilariopsis cylindrus*_272258), NCBI (*Thalassiosira oceanica*_EJK71623.1), Cohen et al. 2018 (*Pseudo-nitzschia granii*) (17), and Moreno et al. 2018 (*Pseudo-nitzschia subcurvata*) (18). **B)** Heatmaps of MCL clusters representing and plastocyanin (cluster_1820) in *Pseudo-nitzschia* and *Fragilariopsis* measured after 24 hours (T1) and 5 days (T5) of incubation under the various iron and temperature treatments. I.E represents ice edge samples, T0 represents *in-situ* samples before any incubations. Each block is one biological replicate measurement. Black-filled up/down pointing triangles represent transcripts that were significantly (glmQLFTest-EdgeR p <0.05) up or down regulated due to warming at T5.

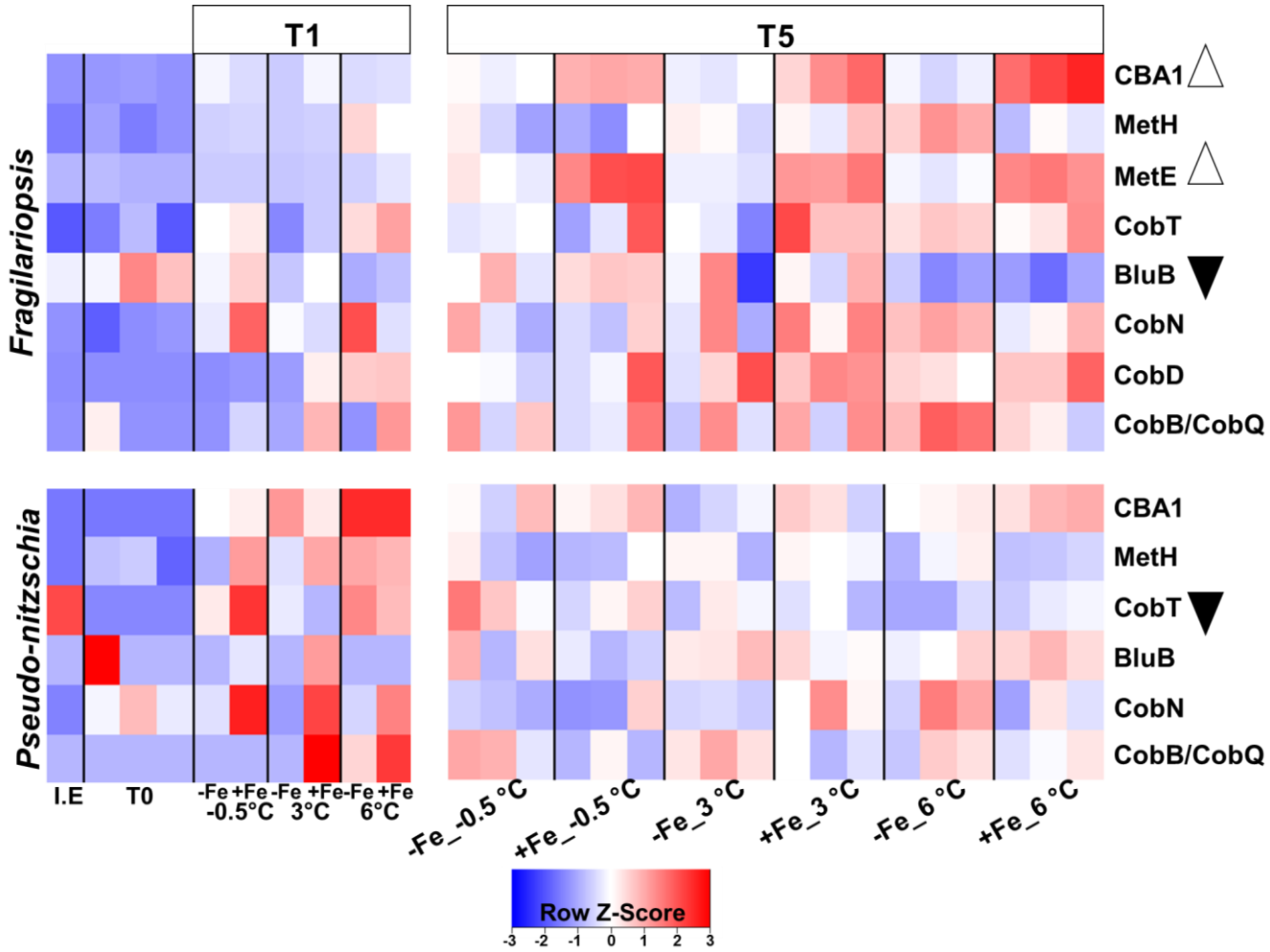

Figure S7 – Heatmaps of MCL clusters involved in B_{12} metabolism in *Fragilariopsis* and *Pseudo-nitzschia* measured after 24 hours or 5 days of incubation under various iron and temperature treatments. I.E represents ice edge samples, T0 represents *in-situ* samples processed in the laboratory before any incubations and each block is one biological replicate measurement. CBA1: cobalamin acquisition protein 1; MetH: cobalamin-requiring methionine synthase; MetE: cobalamin-independent methionine synthase; CobT, CobN: cobaltochelatase; BluB: gene involved in DMB production; CobB/CobQ: cobyrinic acid a,c-diamide synthase/ adenosylcobyric acid synthase. Open triangles represent clusters that were significantly (glmQLFTest-EdgeR $p < 0.05$) up regulated due to iron addition at T5. Black-filled triangles represent clusters that were significantly (glmQLFTest-EdgeR $p < 0.05$) down regulated due to warming at T5.

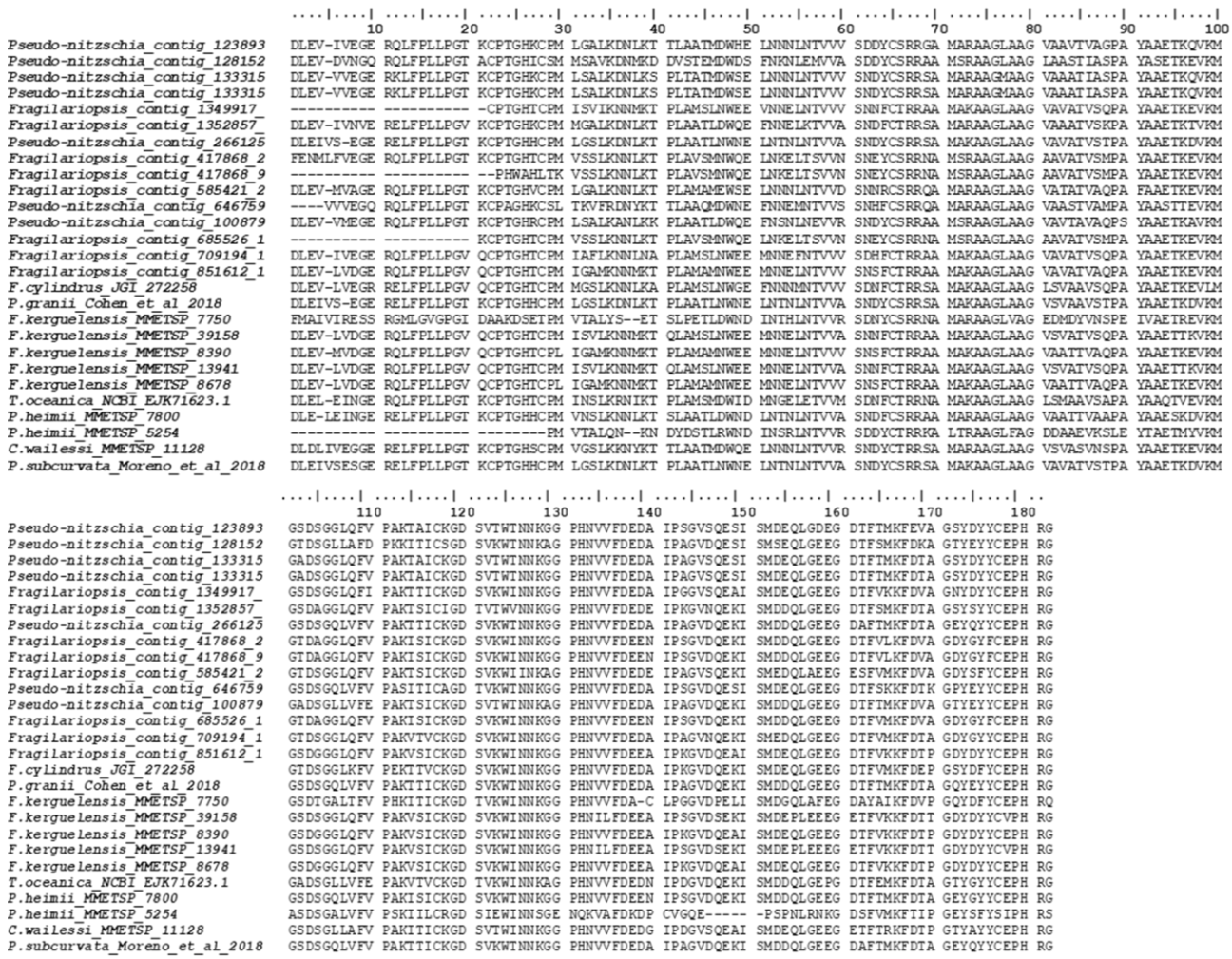

Figure S8 – Alignment of *Pseudo-nitzschia and Fragilariopsis* plastocyanin sequences from our metatranscriptome data and previously identified plastocyanin sequences retrieved from the MMETSP dataset (*Fragilariopsis kerguelensis*_0735*, Pseudo-nitzschia heimii*_1423*, Coscinodiscus wailesii*_1066), JGI (*Fragilariopsis cylindrus*_272258), NCBI (*Thalassiosira oceanica*_EJK71623.1), Cohen et al. 2018 (*Pseudo-nitzschia granii*) (17), and Moreno et al. 2018 (*Pseudo-nitzschia subcurvata*) (16). The alignment was conducted using Clustal Omega in SeaView v5.0. and was used to construct the maximum-likelihood phylogenetic tree for plastocyanin in Fig. S6.

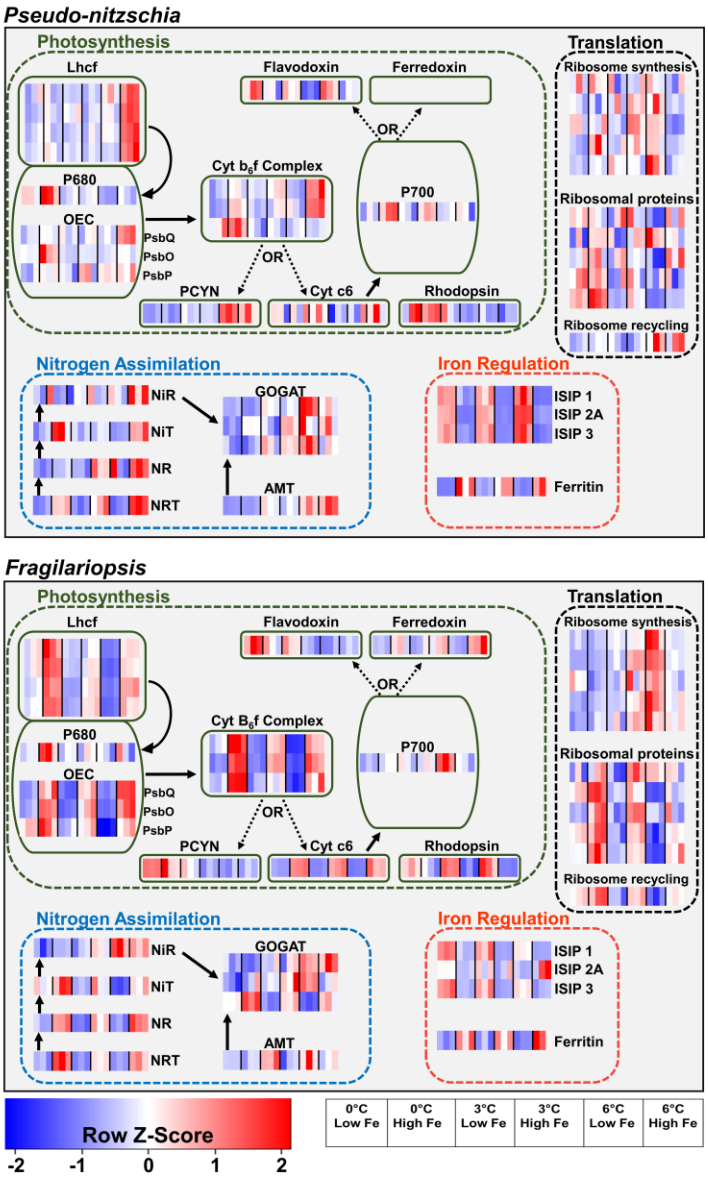

Figure S9 – Schematic representations of *Pseudo-nitzschia* and *Fragilariopsis* cells showing cellular processes, with each process comprised of several protein clusters (MCL clusters). Lhcf = light harvesting complexes-f, OEC = oxygen evolving complex, Cyt B6f complex = cytochrome b_6f complex, PCYN = plastocyanin, Cyt c6 = cytochrome c6, NRT = nitrate transporter, NR = nitrate reductase, NiT = nitrite transporter, NiR = nitrite reductase, AMT = ammonium transporter, GOGAT = glutamine oxoglutarate aminotransferase cycle, ISIP = iron starvation induced protein. Each row in a heatmap represents one Markov cluster (MCL), each column represents a temperature and iron treatment at T5 with each block representing one biological replicate. Heatmaps were constructed using taxon-normalized RPKM values. Empty heatmap placeholders represent clusters found in *Fragilariopsis* but not *Pseudo-nitzschia*. Arrows represent energy/electron flow in photosynthetic light reactions, and steps involved in nitrogen assimilation using nitrate or ammonium.

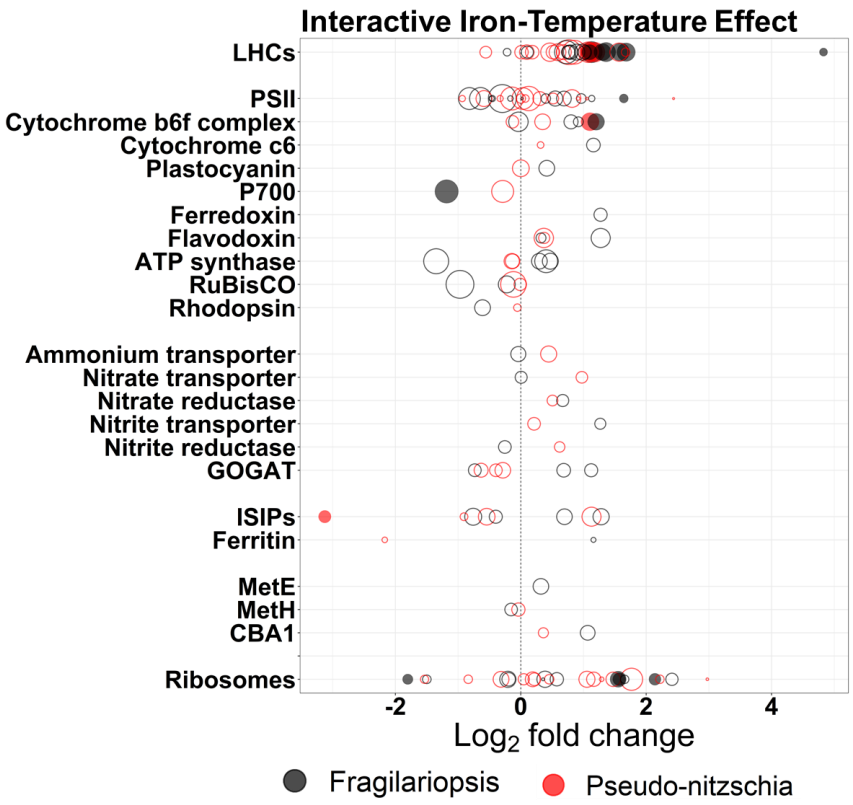

Figure S10. Interactive iron-temperature effect on differential expression (DE) of various clusters. Differential expression was calculated using the quasi-likelihood test (glmQLFTest) in EdgeR and fold-change was calculated for iron-effect at -0.5 °C vs iron-effect at 6 °C treatments at T5. Positive and negative Log2 fold change values represent up and down regulation, respectively. Filled circles are clusters with statistically significant DE (adjusted p-value <0.05). Point size represent total normalized transcript abundance under all iron and temperature treatments.

Table S1 – Projected changes in sea surface temperature and iron availability in the Southern Ocean.

	Temperature	Iron
2100	• +1.5 – 2 °C (27, 28) • +0.5 – 1.5 °C (29)	• + 0.01 nM $m^{-1}_{(upwelled\ water)}$ day^{-1} (28) • -0.002 – -0.006 nM (29)
2300	• +6 °C (28)	• + 0.02 nM $m^{-1}_{(upwelled\ water)}$ day^{-1} (28)

Table S2 – Top ten BLAST-n search results against NCBI's nr/nt database for both *Pseudo-nitzschia* and *Fragilariopsis* 18S rRNA query sequences. UE = Uncultured Eukaryote.

18S rRNA sequence	Matched Taxa	Query Cover	E-Value	% Identity	Accession
Pseudo-nitzschia	*P. subcurvata*	100%	1e-59	100%	KX253952.1
	UE	100%	1e-59	100%	KJ758369.1
	UE	100%	1e-59	100%	KJ758245.1
	Pseudo-nitzschia sp.	100%	1e-59	100%	GU373970.1
	UE	100%	1e-59	100%	AY672814.1
	UE	100%	1e-58	99.23%	HM581774.1
	P. seriata	100%	1e-58	99.23%	GU373969.1
	P. cupsidata	100%	1e-51	96.15%	JN091719.1
	P. lineola	100%	1e-51	96.15%	JN091717.1
	P. turgidula	100%	1e-51	96.15%	FJ222752.1
Fragilariopsis	*F. cylindrus*	100%	1e-59	100%	LC189084.1
	UE	100%	1e-59	100%	KJ758397.1
	UE	100%	1e-59	100%	KJ758375.1
	UE	100%	1e-59	100%	KJ758350.1
	UE	100%	1e-59	100%	KJ758343.1
	UE	100%	1e-59	100%	KJ758332.1
	UE	100%	1e-59	100%	KJ758252.1
	UE	100%	1e-59	100%	KJ758212.1
	UE	100%	1e-59	100%	KJ758191.1
	UE	100%	1e-59	100%	KJ758103.1

**Pseudo-nitzschia* 18S rRNA nucleotide query sequence:
GTCGCACCTACCGATTGAATGGTCCGGTGAAGCCTCGGGATTGTGGCTGGTTTCCTTTATTGGAATCTGACCACGA
GAACCTGTCTAAACCTTATCATTTAGAGGAAGGTGAAGTCGTAACAAGGTTTCC

***Fragilariopsis* 18S rRNA nucleotide query sequence:
GTCGCACCTACCGATTGAATGGTCCGGTGAGGCCTCGGGATTGTGGTTAGTTTCCTTTATTGGAAGTTAGTCGCGA
GAACTTGTCCAAACCTTATCATTTAGAGGAAGGTGAAGTCGTAACAAGGTTTCC

Table S3 – Summary statistics for BLAST-p analyses comparing *Pseudo-nitzschia* peptide sequences from this study and *Pseudo-nitzschia* peptides from various publicly available culture-based transcriptomes. Summary statistics include: number of sequences with better than E = 1E-5 (of the n= 27,420 peptide sequences queried), their match identities and their alignment lengths (1st Quartile, Median, Mean, and 3rd Quartile). Results from *Pseudo-nitzschia subcurvata* are highlighted in bold.

		Match Identities				Alignment Length			
	No. matches	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**
Pseudo-nitzschia arenysensis	23588	56.25	69.58	68.1	81.44	123	195	197	270
Pseudo-nitzschia australis Strain 10249 10AB	21615	51.52	65.9	64.5	78.43	122	193	198	269
Pseudo-nitzschia delicatissima B596	23087	56.02	70	68	81.51	126	198	199	272
*Pseudo-nitzschia delicatissim*a Strain UNC1205	20019	55.98	71.17	68.4	82.58	110	175	181	247
Pseudo-nitzschia fraudulenta Strain WWA7	20896	51.82	67.63	65.8	80.83	117	185	190	261
Pseudo-nitzschia granii	12444	55.28	85.71	75.3	94.82	69	98	110	138
Pseudo-nitzschia heimii Strain UNC1101	22041	53.76	68.28	66.4	80.49	122	193	197	270
Pseudo-nitzschia multiseries	22510	55.47	69.28	67.5	81.18	108	173	181	249
Pseudo-nitzschia pungens	22333	51.52	65.53	64.5	78.28	116	188	193	268
Pseudo-nitzschia pungens cf. cingulata	21727	50.41	64.67	63.8	77.83	121	193	197	271
***Pseudo-nitzschia subcurvata* (Moreno et al 2017)**	**24647**	**93.47**	**98.44**	**90.4**	**99.57**	**119**	**179**	**188**	**255**

Table S4 – BLAST-p search results for sequences encoding domoic acid biosynthesis proteins (DabA, B, C, D) retrieved from GenBank, against all ORFs from this study. No matches were found for DabA and DabB encoding genes and no significant eukaryotic matches were found for DabC encoding genes. Score and e-values were calculated using the Blosum62 similarity matrix. E-value $1e^{-30}$ was used as the cut-off for DabD results. These DabD results are further explored in Table S5.

Query Gene	Matched Sequence (ORF ID)	Score	E-value	Taxa	Hypothesized annotation
DAB-A AYD91073.1	-	-	-	-	-
DAB-B AYD91072.1	-	-	-	-	-
DAB-C AYD91075.1	**contig_254441_113_1024_+**	63	6e-010	Flavobacteria	2OG-Fe(II) oxygenase family
	contig_1174518_67_906_+	61	6e-010	Alteromonadales	2OG-Fe(II) oxygenase family
	contig_732674_1_798_-	56	9e-008	Alteromonadales	2OG-Fe(II) oxygenase family
	contig_633551_118_1041_+	56	9e-008	Flavobacteria	2OG-Fe(II) oxygenase family
	contig_627399_1_787_-	54	5e-007	Cyanobacteria	2OG-Fe(II) oxygenase family
	contig_1862_1_732_-	54	5e-007	Other Bacteria	2OG-Fe(II) oxygenase family
	contig_660976_1_1043_+	52	2e-006	Other Stramenopiles	2OG-Fe(II) oxygenase family
	contig_253357_26_973_+	52	2e-006	Flavobacteria	2OG-Fe(II) oxygenase family
DAB-D AYD91074.1	**contig_596040_24_938_-**	231	9e-061	Pseudo-nitzschia	Cyt P450, CYP4/CYP19/CYP26
	contig_625510_955_1881_-	225	6e-059	Pseudo-nitzschia	Cyt P450, CYP4/CYP19/CYP26
	contig_82244_85_1023_+	156	5e-038	Fragilariopsis	Cyt P450, CYP4/CYP19/CYP26
	contig_596040_947_1648_-	152	9e-037	Fragilariopsis	Cyt P450, CYP4X1
	contig_82244_1395_1877_+	145	7e-035	Fragilariopsis	Cyt P450
	contig_680152_23_630_-	132	6e-031	Cyanobacteria	Cyt P450
	contig_1109498_1_752_-	130	3e-030	Other Diatom	Cyt P450

412 **Table S5 –** Reciprocal BLAST-p search results for ORFs with similarity to DabD-encoding genes
413 (Table S4) against NCBI's nr database.

ORF ID	Top four BLAST-p hits	Taxa	Query Cover	E-Value	% Identity	Accession
contig_596040_24_938_-	- Cytochrome P450	*F. cylindrus*	99%	2e-133	64.62%	OEU10111.1
	- Unnamed protein	*P. multistriata*	99%	4e-71	42.95%	VEU44693.1
	- DabD	*P. multiseries*	98%	4e-70	42.43%	AYD91074.1
	- Alkane-1-monooxygenaze	*F. solaris*	96%	2e-61	39.86%	GAX28661.1
contig_625510_955_1881_-	- Cytochrome P450	*F. cylindrus*	94%	1e-111	56.48%	OEU10111.1
	- DabD	*P. multiseries*	95%	3e-67	41.81%	AYD91074.1
	- Unnamed protein	*P. multistriata*	95%	3e-63	41.14%	VEU44693.1
	- Alkane-1-monooxygenaze	*F. solaris*	94%	6e-58	40.61%	GAX28661.1
contig_82244_85_1023_+	- Cytochrome P450	*F. cylindrus*	99%	0.0	96.43%	OEU10111.1
	- Alkane-1-monooxygenaze	*F. solaris*	88%	5e-46	35.21%	GAX23170.1
	- Alkane-1-monooxygenaze	*F. solaris*	88%	2e-45	35.92%	GAX28661.1
	- DabD	*P. multiseries*	93%	3e-41	32.00%	AYD91074.1
contig_596040_947_1648_-	- Cytochrome P450	*F. cylindrus*	93%	1e-114	77.06%	OEU10111.1
	- Alkane-1-monooxygenaze	*F. solaris*	84%	4e-41	40.8%	GAX28661.1
	- Alkane-1-monooxygenaze	*F. solaris*	84%	2e-40	40.5%	GAX23170.1
	- DabD	*P. multiseries*	84%	8e-40	37.81%	AYD91074.1
contig_82244_1395_1877_+	- Cytochrome P450	*F. cylindrus*	100%	4e-110	99.37%	OEU10111.1
	- Hypothetical protein	*T. oceanica*	98%	3e-43	49.68%	EJK45228.1
	- DabD	*P. multiseries*	99%	1e-38	47.47%	AYD91074.1
	- Unnamed protein	*P. multistriata*	98%	2e-37	45.86%	VEU44693.1
contig_680152_23_630_-	- TPA: cytochrome P450	Porticoccaceae	100%	5e-130	86.07%	HAZ79708.1
	- Hypothetical protein	Porticoccaceae	100%	2e-128	86.07%	MAY69286.1
	- Cytochrome P450	Porticoccaceae	100%	7e-102	69.65%	WP_155531439.1
	- Hypothetical protein	SAR92	99%	1e-101	69.50%	KRP17789.1
contig_1109498_1_752_-	- Cytochrome P450	*F. cylindrus*	92%	7e-58	43.25%	OEU10111.1
	- DabD	*P. multiseries*	91%	1e-32	29.34%	AYD91074.1
	- Unnamed protein	*P. multistriata*	91%	6e-31	29.75%	VEU44693.1
	- Alkane-1-monooxygenaze	*F. solaris*	87%	1e-24	30.38%	GAX28661.1

414

415

Table S6 – Protein ID (PID) assignments for ORFs in the various *Pseudo-nitzschia* spp. and *Fragilariopsis* spp. light harvesting complex (LHC) clusters. PIDs were assigned by performing a BLAST-p search against *Pseudo-nitzschia multiseries* (CLN-47) and *Fragilariopsis cylindrus* (CCMP 1102). LHC groups were assigned based on previous diatom LHC classifications in Mock et al. 2017 (Supp.Info.11) and Hippmann et al. 2017 (14, 15).

Cluster	*Pseudo-nitzschia* PID	*Fragilariopsis* PID	LHC Group
clust_1044	258347, 261276	169285, 205888	**Lhcf_III**
clust_1084	247179, 263274	195639, 261294	**Lhcr**
clust_1194	257565, 303201, 304112, 306047	195777	**Lhcf_I**
clust_1236	41763, 197371, 302398, 310027	170761, 174589, 269349, 269868, 271557	**Lhcf_I**
clust_1787	238335, 257821	188478, 271659	**Lhcx**
clust_1938	318557	270184	**Lhcr**
clust_2256	66239, 238335	269313, 272562	**Lhcx**
clust_2549	307175	273003, 271559	**Lhcr**
clust_3219	301726	270606	**Lhcr**
clust_3282	307174	271561, 273005	**Lhcf_II**
clust_332	257565, 306047, 306447	267329, 271330, 271332	**Lhcf_I**
clust_3363	264176	260998	**Lhcr**
clust_3551	325841	210115, 213124	**Lhcr**
clust_402	191001, 300768, 303201, 304112, 305325, 307376	267576, 267702, 267837, 268624, 269543, 269567	**Lhcf_I**
clust_646	14959, 178030, 234364, 318210	177731, 187698, 195639, 270940	**Lhcz**
clust_712	66239, 264022, 307174	218498, 271659, 272024	**Lhcx**
clust_80	191001, 255698, 300768, 300769, 303058, 303201, 305325, 305720, 307376	143190, 174589, 207327, 267576, 271931, 268626	**Lhcf_I**

Table S7 – Summary statistics for both assembled contigs and predicted ORFs for all replicates and treatments combined.

Assembly	Assembled Contigs	Predicted ORFs from contigs
# contigs (≥ 0 bp)	1315493	2265230
# contigs (≥ 1000 bp)	233450	26705
# contigs (≥ 5000 bp)	1445	0
# contigs (≥ 10000 bp)	106	0
# contigs (≥ 25000 bp)	3	0
# contigs (≥ 50000 bp)	0	0
Total length (≥ 0 bp)	937746248	689493243
Total length (≥1000 bp)	359162487	29719929
Total length (≥ 5000 bp)	9769226	0
Total length (≥ 10000 bp)	1394315	0
Total length (≥ 25000 bp)	84630	0
Total length (≥ 50000 bp)	0	0
# contigs	795370	384770
Largest contig	33111	3999
Total length	754213377	276614610
GC (%)	43.32	45.72
N50	969	732
N75	710	600
L50	251690	153883
L75	480396	258428
# N's per 100 kbp	91.56	4.23

 # contigs (≥ x bp): total number of contigs of length ≥ x bp.
Total length (≥ x bp): total number of bases in contigs of length ≥ x bp.
 # contigs: total number of contigs in the assembly for contigs size ≥ 500 bp.
Largest contig is the length of the longest contigs in the assembly.
Total length is the total number of bases in the assembly for using contigs size ≥ 500 bp.
GC (%): total number of G and C nucleotides in the assembly, divided by the total length of the assembly.
N50: length for which the collection of all contigs of that length or longer covers at least half (50%) the total base content of the Assembly. It serves as a median value for assessing whether the Assembly is balanced towards longer contigs (higher N50) or shorter contigs (lower N50). N75 is used for the same purpose but the length is set at 75% of total base content instead of 50%.
L50: number of contigs equal to or longer than the N50 length. In other words, L50, is the minimal number of contigs that contain half the total base content of the Assembly. L75 is used for the same purpose in reference to the N75 length.
 # N's per 100 kbp: average number of uncalled bases per 100,000 assembly bases.

438 **Table S8 –** Total number of raw and trimmed reads, mRNA and rRNA contribution to the trimmed
439 reads, number of mapped mRNA reads, and total number of predicted ORFs for all individual
440 replicates and treatments.

Treatment	Raw Reads	Trimmed Reads	mRNA	rRNA	% rRNA	Mapped Reads	% Mapped Reads	# ORFs
Ice Edge	28492536	27232733	25443952	1788781	6.6%	18226962	71.6%	826190
T0_A	45575920	38947343	33358200	5589143	14.4%	23296503	69.8%	852002
T0_B	19375404	17965033	15308130	2656903	14.8%	10733771	70.1%	577573
T0_C	46579930	38952747	28025278	10927469	28.1%	20712096	73.9%	638446
T1_-Fe_-0.5 °C	21205314	19401307	17366099	2035208	10.5%	12812358	73.8%	576148
T1_+Fe_-0.5 °C	38041776	35740347	32575458	3164889	8.9%	23588683	72.4%	1025062
T1_-Fe_3 °C	35349994	32732641	28188283	4544358	13.9%	20572627	73.0%	916171
T1_+Fe_3 °C	34477352	33059336	29594763	3464573	10.5%	21540022	72.8%	1008607
T1_-Fe_6 °C	32027512	30406112	26596077	3810035	12.5%	18668646	70.2%	887919
T1_+Fe_6 °C	29129438	27409840	23359243	4050597	14.8%	17033876	72.9%	825869
T5_-Fe_-0.5 °C_A	37671290	33464510	28737672	4726838	14.1%	20375592	70.9%	986301
T5_-Fe_-0.5 °C_B	30654564	29139435	26199569	2939866	10.1%	18903422	72.2%	1031680
T5_-Fe_-0.5 °C_C	40290894	36930817	32542651	4388166	11.9%	23560702	72.4%	1128714
T5_+Fe_-0.5 °C_A	39948682	38527552	34406982	4120570	10.7%	25220490	73.3%	1080944
T5_+Fe_-0.5 °C_B	38691956	35399483	29137122	6262361	17.7%	21608015	74.2%	1003994
T5_+Fe_-0.5 °C_C	36287660	34523817	30608808	3915009	11.3%	22940050	75.0%	1013086
T5_-Fe_3 °C_A	34696294	33814638	30373744	3440894	10.2%	22017380	72.5%	1161828
T5_-Fe_3 °C_B	31585290	30088753	27450391	2638362	8.8%	19637249	71.5%	1084754
T5_-Fe_3 °C_C	35546072	32177978	27306946	4871032	15.1%	18999285	69.6%	853823
T5_+Fe_3 °C_A	28153060	26415130	24099272	2315858	8.8%	17773329	73.8%	803545
T5_+Fe_3 °C_B	46637782	41924346	34537148	7387198	17.6%	26251536	76.0%	975259
T5_+Fe_3 °C_C	22038026	20474922	17797508	2677414	13.1%	13314623	74.8%	636362
T5_-Fe_6 °C_A	34470852	32034859	28986089	3048770	9.5%	21576040	74.4%	954425
T5_-Fe_6 °C_B	28342516	27016685	22841677	4175008	15.5%	16929366	74.1%	692317
T5_-Fe_6 °C_C	35887516	35118299	28397352	6720947	19.1%	20889743	73.6%	971567
T5_+Fe_6 °C_A	31556152	29705369	26100052	3605317	12.1%	19817298	75.9%	733987
T5_+Fe_6 °C_B	40514850	38946408	35004303	3942105	10.1%	26683763	76.2%	971080
T5_+Fe_6 °C_C	24900074	23356091	19248203	4107888	17.6%	14118461	73.4%	647783

441

442

Table S9 – Number of reads assigned to *Fragilariopsis* and *Pseudo-nitzschia* for all treatments and replicates.

Treatment	*Fragilariopsis*	*Pseudo-nitzschia*
Ice Edge	604153	149926
T0_A	792326	265315
T0_B	324368	87393
T0_C	444067	132386
T1_-Fe_-0.5 °C	525883	141403
T1_+Fe_-0.5 °C	1389023	422961
T1_-Fe_3 °C	1032805	285845
T1_+Fe_3 °C	1298385	452981
T1_-Fe_3 °C	1270679	501788
T1_+Fe_3 °C	896978	305966
T5_-Fe_-0.5 °C_A	2366719	872030
T5_-Fe_-0.5 °C_B	1783721	642148
T5_-Fe_-0.5 °C_C	2192675	875719
T5_+Fe_-0.5 °C_A	2618615	833378
T5_+Fe_-0.5 °C_B	2381675	694578
T5_+Fe_-0.5 °C_C	3172422	1140824
T5_-Fe_3 °C_A	2433791	1503958
T5_-Fe_3 °C_B	1871392	1150238
T5_-Fe_3 °C_C	1348818	715756
T5_+Fe_3 °C_A	2531661	1003347
T5_+Fe_3 °C_B	3390565	1619939
T5_+Fe_3 °C_C	2199219	909139
T5_-Fe_6 °C_A	2406422	2404759
T5_-Fe_6 °C_B	2216364	2339347
T5_-Fe_6 °C_C	3528660	2814964
T5_+Fe_6 °C_A	2772000	1577368
T5_+Fe_6 °C_B	4251421	2819891
T5_+Fe_6 °C_C	2445033	1409944

1. E. M. Bertrand, *et al.*, Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge. *Proceedings of the National Academy of Sciences* **112**, 9938–9943 (2015).

2. R. Schmieder, Y. W. Lim, R. Edwards, Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).

3. M. Rho, H. Tang, Y. Ye, FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Research* **38**, e191–e191 (2010).

4. S. Podell, T. Gaasterland, DarkHose: a new method for genome-wide prediction of horizontal gene transfer. *Genome Biol* **8**, R16 (2007).

5. A. J. Enright, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**, 1575–1584 (2002).

6. L. A. Amaral-Zettler, E. A. McCliment, H. W. Ducklow, S. M. Huse, A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE* **4**, e6372 (2009).

7. E. Bolyen, *et al.*, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).

8. M. Martin, Cutadapt removes adapter sequences from high-throughput reads. *EMBnet.journal* **17**, 3 (2011).

9. B. J. Callahan, *et al.*, DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13** (2016).

10. N. A. Bokulich, *et al.*, Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).

11. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12** (2011).

12. L. Guillou, *et al.*, The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **41**, D597–D604 (2012).

13. J. K. Brunson, *et al.*, Biosynthesis of the neurotoxin domoic acid in a bloom-forming diatom. *Science* **361**, 1356–1358 (2018).

14. T. Mock, *et al.*, Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **541**, 536–540 (2017).

15. A. A. Hippmann, *et al.*, Contrasting effects of copper limitation on the photosynthetic apparatus in two strains of the open ocean diatom Thalassiosira oceanica. *PLoS ONE* **12**, e0181753 (2017).

16. C. M. Moreno, *et al.*, Examination of gene repertoires and physiological responses to iron and light limitation in Southern Ocean diatoms. *Polar Biol* **41**, 679–696 (2018).

17. N. R. Cohen, *et al.*, Transcriptomic and proteomic responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation: Intracellular processes of an iron-limited diatom. *Environ Microbiol* **20**, 3109–3126 (2018).

18. T. Parsons, Y. Maita, C. Lalli, *A manual of chemical and biological methods for seawater analysis* (New York: Pergamon Press, 1984).

19. F. Koroleff, *Determination of nutrients* (New York: Verlag Chemie, 1983).

20. R. C. Dugdale, F. P. Wilkerson, The use of ^{15}N to measure nitrogen uptake in eutrophic oceans; experimental considerations. *Limnol. Oceanogr.* **31**, 673–689 (1986).

21. T. Hama, *et al.*, Measurement of photosynthetic production of a marine phytoplankton population using a stable 13C isotope. *Mar. Biol.* **73**, 31–36 (1983).

22. S. E. Baer, *et al.*, Seasonal nitrogen uptake and regeneration in the western coastal Arctic. *Limnology and Oceanography* **62**, 2463–2479 (2017).

23. A. O. Tatters, *et al.*, Interactive effects of temperature, CO_2 and nitrogen source on a coastal California diatom assemblage. *Journal of Plankton Research* **40**, 151–164 (2018).

24. C. Tomas, *Identifying Marine Phytoplankton* (Academic Press, San Diego, 1997).

25. F. I. Scott, H. J. Marchant, Eds., *Antarctic Marine Protists* (ABRS, Canberra and Australian Antarctic Division, 2005).

26. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).

27. P. W. Boyd, S. T. Lennartz, D. M. Glover, S. C. Doney, Biological ramifications of climate-change-mediated oceanic multi-stressors. *Nature Clim Change* **5**, 71–79 (2015).

28. J. K. Moore, *et al.*, Sustained climate warming drives declining marine biological productivity. *Science* **359**, 1139–1143 (2018).

29. P. W. Boyd, *et al.*, Climate-mediated changes to mixed-layer properties in the Southern Ocean: assessing the phytoplankton response. 18 (2008).