Supplementary Information

This supplementary information file contains the following:

- **Supplementary Methods** *(additional details)*
    - *Culturing experiments*
    - *Protein extraction, digestion, and mass spectrometry analyses*
    - *RNA extraction, Illumina sequencing, and annotation*
    - *Ross Sea Phaeocystis bloom sample collection and protein extraction and analysis*
    - *Field pigment analysis*
    - *Data availability*
- **Supplementary Tables 1-3**
- **Supplementary Fig. 1-10**
- **Supplementary References** *(to accompany methods)*
- **Supplementary Datasets 1 and 2** *(included as separate excel files)*
    - **Supplementary Data 1.** Complete proteomic results from *P. antarctica* culture iron limitation experiments.
    - **Supplementary Data 2.** Metaproteomic and metatranscriptomic results from Ross Sea 2005 sample using the three databases.

Supplementary Methods

Culturing experiments

Two strains of *Phaeocystis antarctica* (treated with Provasoli's antibiotic concentrated solution with 250 μg mL^{-1} cefotaxime and 500 μg mL^{-1} carbopenicillin), CCMP 1871 and CCMP 1374 (Provasoli-Guillard National Center for Culture of Marine Phytoplankton), and a Ross Sea centric diatom isolate *Chaetoceros* sp. RS-19 (collected by M. Dennett at 76.5° S, 177.1° W in December 1997 and isolated by D. Moran) were grown in F/2 media with a trace metal stock (minus $FeCl_3$) according to Sunda and Huntsman (Sunda and Huntsman, 2003; 1995), using a modified 10 μM EDTA concentration, and an oligotrophic seawater base. Strains were chosen because they were culturable representatives from two distinct regions in the Southern Ocean. The oligotrophic seawater base was collected from surface South Atlantic Ocean water as part of the 2007 CoFeMUG cruise - station 13 where iron concentration was measured at 0.05 nM (Noble et al., 2012). Water was 0.2 μm filtered using trace metal clean techniques, and microwave sterilized. Filter-sterilized nutrients and vitamins (882 μM $NaNO_3$, 42 μM NaH_2PO_4, 106 μM Na_2SiO_3, 0.4 nM Biotin, 60 nM Thiamine HCl, 0.074 nM B_{12}) were run through a Chelex-100 column with pre-cleaned chelex (Price et al., 1989), and were then added to media. All cultures were maintained in acid-washed polycarbonate vessels prepared in a class 100 clean room facility.

Semi-continuous batch cultures were grown at 4 °C under 200 μmol photons m^{-2} s^{-1} continuous light. Each strain was acclimated to growth on one of six total added iron (Fe_{Total}) concentrations: 0 nM, 1 nM, 3 nM, 10 nM, 30 nM, and 100 nM, plus the seawater blank of 0.05nM Fe described above. The concentration of dissolved inorganic iron within each treatment was calculated using the empirically measured literature values collected under similar conditions by Sunda and Huntsman (Sunda and Huntsman, 2003) with a Fe'/Fe_{Total} = 0.039 at 10 μM EDTA,

under 500 µE illuminated conditions, and at 10 ºC temperature, for Fe' concentrations of 2 pM, 41 pM, 120 pM, 740 pM, 1200 pM, and 3900 pM, where Fe' refers to the sum of inorganic ferric species not bound to EDTA. Cultures were grown in 28 mL polycarbonate Nalgene tubes and monitored using a Turner Designs TD 700 Fluorometer. Growth rates were calculated based on RFUs (relative fluorescence units); cultures were considered acclimated to growth after a minimum of three transfers for each strain in each condition.

Acclimated *P. antarctica* cultures were transferred into 250 mL polycarbonate bottles to begin each proteome and transcriptome experiment. Every 1-2 days subsamples were collected in 5 mL 13x100 mm borosilicate tubes to measure RFUs and cell counts in the treatments. Cell counts were conducted using a Palmer-Maloney counting chamber and a Zeiss Axio Plan microscope on 400x magnification; cell numbers were used to determine the final growth rate of each strain/treatment. During mid-to-late exponential phase (time-of-harvest), cell size was determined for both strains (n=20 cells were counted for each strain), calculated using the Zeiss 4.8.2 software and a calibrated scale bar. The number of cells in colonies (versus as single cells) was determined for strain 1871 only. Briefly, counts (number of cells associated with colonies versus unassociated) were averaged from 10 fields of view at five distinct time points (50 fields of view total).

For strain 1871, 200 mL of culture in mid-exponential phase was centrifuged at 12,428 x g for 20 min at 4 ºC in trace metal clean centrifuge bottles to form pellets. For strain 1374, 200 mL of culture in late-exponential phase was pelleted as above. Each pellet was resuspended in 1-2 mL of fresh F/2 –$FeCl_3$ media and transferred into a 1.5 mL eppendorf tube. The tubes were then spun for 10 min at 4 ºC and 6,700 x g. The supernatant was removed and the pellets were frozen at -80 ºC.

Protein extraction, digestion, and mass spectrometry analyses

Cell pellets reserved for protein extractions were thawed once and then freeze/thawed two more times using liquid nitrogen (one pellet per treatment, two strains and 6 iron treatments for a total of 12 proteome samples). Then, 800 µL of B-PER reagent (Thermo Scientific) was added to each pellet, followed by one freeze/thaw cycle, and a final freeze step. Frozen material was sonicated (at an output of 2.5) for 4min on ice, and left on ice for 10min before a second round of sonication. Sonicated samples were centrifuged at 4 ºC for 20 min at 14,100 x g. The supernatant was removed and added to a new, methanol-clean tube. To precipitate the protein, 1600 µL 50:50 methanol:acetone with HCl (0.5 mM) was added to the 400 µL in each. Tubes were precipitated at -20 ºC overnight. The next day, the tubes were spun down for 30 min (4 ºC at 14,1000 x g), the supernatant was removed, and the pellets were dried for 5min in the speed vac. Pellets were resuspended in B-PER according to pellet size and allowed to resuspend at 4 ºC for 1 h. Protein was quantified using the BioRad DC Protein Assay Quantification Kit (BioRad Inc., Hercules CA), using BSA as the standard. In all cases there was plenty of material for analysis, with total protein yields typically 64-574 µg total protein for Strain 1374, and 29-217 µg for Strain 1871 across the treatments (only 1 µg total protein was injected for each proteomic analysis).

All extracted proteins were purified from detergent, reduced, alkylated and trypsin digested while embedded within a polyacrylamide tube gel, modified from a previously published method (Lu and Zhu, 2005). A gel premix was made by combining 1M Tris HCL (pH 7.5) and 40% Bis-acrylimide L 29:1 (Acros Organics) at a ratio of 1:3. The premix (103 µL) was combined with an extracted protein sample (35 µg-200 µg), TE, 7 µl 1% APS, and 3 µl of TEMED (Acros Organics) to a final volume of 200 µL. After 1h of polymerization at RT 200 µL of gel fix solution (50% ETOH, 10% acetic acid in LC/MS grade water) was added to the top of the gel and incubated at

RT for 20 min. Liquid was then removed and the tube gel was transferred into a new 1.5 mL microtube containing 1.2 mL of gel fix solution then incubated at RT, 350 rpm in a Thermomixer R (Eppendorf) for 1 h. Gel fix solution was then removed and replaced with 1.2 mL destain solution (50% MeOH, 10% acetic acid in H_2O) and incubated at 350 rpm, RT for 2 h. Liquid was then removed, the gel cut up into 1mm cubes and then added back to tubes containing 1mL of 50:50 acetonitrile:25 mM ammonium bicarbonate (ambic) incubated for 1 h, 350 rpm at RT. Liquid was removed and replaced with fresh 50:50 acetonitrile:ambic and incubated at 16 °C 350 rpm overnight. The above step was repeated for 1h the following morning. Gel pieces were then dehydrated twice in 800 µL of acetonitrile for 10 min at RT and dried for 10 min in a ThermoSavant DNA110 speedvac after removing solvent. 600 µL of 10 mM DTT in 25 mM ambic was added to reduce proteins incubating at 56 °C, 350 rpm for 1 h. Unabsorbed DTT solution was then removed with volume measured. Gel pieces were washed with 25 mM ambic and 600 µL of 55 mM iodacetamide was added to alkylate proteins at RT, 350 rpm for 1 h. Gel cubes were then washed with 1 mL ambic for 20 min, 350 rpm at RT. Acetonitrile dehydrations and speedvac drying were repeated as above. Trypsin (Promega #V5280) was added in appropriate volume of 25 mM ambic to rehydrate and submerse gel pieces at a concentration of 1:20 µg trypsin:protein. Proteins were digested overnight at 350 rpm 37 °C. Unabsorbed solution was removed and transferred to a new tube. 50 µL of peptide extraction buffer (50% acetonitrile, 5% formic acid in water) was added to gels, incubated for 20 min at RT then centrifuged at 14,100 x g for 2 min. Supernatant was collected and combined with unabsorbed solution. The above peptide extraction step was repeated combining all supernatants. Combined protein extracts were centrifuged at 14,100 x g for 20 min, supernatants transferred into a new tube, and dehydrated down to approximately 10 µL-20 µL in the speedvac. Concentrated peptides were then diluted in 2%

acetonitrile and 0.1% formic acid in water for storage until analysis. All water used in the tube gel digestion protocol was LC/MS grade, and all plastic microtubes were ethanol rinsed and dried prior to use.

Digested protein extracts were analyzed by liquid chromatography-mass spectrometry (LC-MS) (Michrom Advance HPLC coupled to a Thermo Scientific Q-Exactive Orbitrap mass spectrometer with a Michrom Advance CaptiveSpray source). Each sample (1 µg protein measured before tryptic digestion) was concentrated onto a trap column (0.3 x 10 mm ID, 3 µm particle size, 200 Å pore size, SGE Protocol C18G) and rinsed with 150 µL 0.1% formic acid, 5% acetonitrile (ACN), and 94.9% water before gradient elution through a reverse phase C18 column (0.15 x 150mm ID, 3µm particle size, 200 Å pore size, SGE Protocol C18G) at a flow rate of 1 µL min^{-1}. The chromatography consisted of a nonlinear 210 min gradient from 5% to 95% buffer B, where A was 0.1% formic acid in water and B was 0.1% formic acid in ACN (all solvents were Fisher Optima grade). The mass spectrometer was set to perform MS/MS on the top 15 ions using data-dependent settings (dynamic exclusion 30 s, excluding unassigned and singly charged ions), and ions were monitored over a range of 380-2000 m/z.

Protein identifications were generated using Proteome Discoverer 1.4 (Thermo Scientific) using the translated transcriptomes for *P. antarctica* strain 1871 and strain 1374, respectively analyzed from a range of iron concentrations from this study (see transcriptome methods). Spectral scores were counted using Scaffold 4.0 (Proteome Software Inc.) with a protein false discovery rate (FDR) of 1.0%, a minimum peptide score of 2, and a peptide probability threshold of 95%. Subsequent proteomic analyses were conducted using the R software package after filtering the data to include a spectral abundance score of "3" for at least one replicate in at least one treatment. The R package "FactoMineR"(Lê et al., 2008) was used for the PCA analysis; for heatmaps, the

package “gplots” was used (Warnes et al., 2009). Proteomic samples taken from each laboratory condition were not pooled downstream as part of the analyses; replicates shown for each treatment are technical replicates.

RNA extraction, Illumina sequencing, and annotation

For *P. antarctica* cultures total RNA was isolated from cell pellets (one pellet per treatment, two strains and three iron concentrations for six samples) following the TRIzol Reagent (Life Technologies, manufacturer’s protocol). RNeasy Mini kit (Qiagen) was used for RNA cleanup, and DNase I (Qiagen) treatment was applied in solution and on-column for removing genomic DNA during RNA cleanup. Purified total RNA integrity was qualitatively evaluated using an Agilent Technologies 2100 Bioanalyzer and quantitated by Qubit. Libraries were constructed using a TruSeq RNA Sample Preparation Kit V2 (IlluminaTM) (0.8 μg of purified total RNA was the input), following the manufacturer’s TruSeq RNA Sample Preparation Guide. Final libraries were purified and size-selected by Agencourt AMPure XP kit (Beckman Coulter Genomics). Library quality was evaluated with an Agilent Technologies 2100 Bioanalyzer using a High Sensitivity DNA chip; mean library size was around 500 bp. Purified material was subsequently subjected to paired-end sequencing.

For metatranscriptomes, RNA was extracted from the frozen cell pellets using the TRIzol reagent manufacturer’s protocol (Thermo Fisher Scientific). 6-8 μg of total RNA was obtained per pellet of biomass. Two RNA-seq libraries were constructed; one from rRNA depleted mRNA and a second from polyA enriched mRNA. In the case of rRNA depleted mRNA, an RNeasy Mini Kit was applied (Qiagen), ribosomal RNA was removed from 2.5 μg of total RNA with Ribo-Zero Magnetic kits, employing a mix of plant, bacterial, and human/mouse/rat formulations in a ratio

of 2:1:1 (Epicentre). The resulting mRNA enrichment was purified using an Agencourt RNAClean XP kit (Beckman Coulter) and 5 ng of rRNA depleted RNA was used for library construction using ScriptSeq V2 RNA-seq Library Preparation kit (Epicentre). The library was then subjected to paired-end sequencing via Illumina Hiseq, resulting in 35M reads with an average of 150 bp per read.

For all samples, reads were trimmed for quality and filtered to remove primers, adaptors and rRNA sequences using Ribopicker v.0.4.3 (Schmieder et al., 2012). CLC Assembly Cell (CLCbio; version 3.22.55708) was used to assemble contigs, open reading frames (ORFs) were predicted from the assembled contigs using FragGeneScan, and additional rRNA sequences were identified and removed (Rho et al., 2010). The remaining ORFs were annotated de novo for function using a reference database as described below. Phylogeny was assigned to each ORF using a Lineage Probability Index (LPI) (Podell and Gaasterland, 2007). LPI was adapted for this study to prevent incorrect taxonomic assignment of highly conserved sequences and sequences similar to contaminants in the reference dataset (e.g. bacterial contaminant sequences in new eukaryotic transcriptome references, see below). ORFs identified as organelle- derived were classified separately from bacterial and nuclear mRNA. Group normalized RPKM values were calculated to compare abundances of different ORFs for each phylogenetic grouping (e.g. reads per kilobase per million diatom reads mapped = diatom RPKM). For *P. antarctica* experiments, ORFs classified as "Haptophyta" using the LPI taxonomic string were retained for downstream analyses. Analysis of sequence counts ("ASC") was used to assign normalized fold change and determine which ORFs were significantly differentially expressed in pairwise comparisons between treatments, and reported an adjusted p-value rather than applying a significance threshold.

ASC approach offers a robust analysis of differential gene expression data for non-replicated samples (Wu et al., 2010).

The custom database applied here was comprised of peptides obtained from KEGG, GenBank, JGI, ENSEMBL, CAMERA, and various other repositories. Altogether, version 1.075 of the database consists of 24,509,327 peptides from 19,962 viral, 230 archaeal, 4910 bacterial, and 894 eukaryotic taxa. Importantly for analyses reported in this study, the database also contains new data from the Marine Microbial Eukaryotic Transcriptome Sequencing Project (http://marinemicroeukaryotes.org/), which is represented by 8,807,335 peptides from 411 taxa. ORFs identified in these metatranscriptomic analyses were mapped to the *Oceanospirillaceae* ASP10-02a, as well as other draft genomes recovered from metagenomics sequencing of the microbial community associated with a *P. antarctica* bloom event in the Amundsen Sea polynya (West Antarctica)(Bertrand et al., 2015; Delmont et al., 2015). ORFs were associated with a particular metagenome bin based on the best scoring nucleotide alignment using BWA-MEM http://arxiv.org/abs/1303.3997 with default parameters (Li, 2013). For annotations based on HMMER or BLASTP, we use an e-value threshold of 1e-3.

Ross Sea Phaeocystis bloom sample collection and protein extraction and analysis

The metaproteome sample was collected in the Ross Sea (170.76° E, 76.82° S) during the CORSACS expedition (Controls on Ross Sea Algal Community Structure) on December 30, 2005 (near pigment station 137; http://www.bco-dmo.org/dataset-deployment/453377) (Saito et al., 2010; Sedwick et al., 2011). Surface water was concentrated via a plankton net tow (20µm mesh), gently decanted of extra seawater, then split into multiple replicate cryovials and frozen in RNAlater at -80 °C, for metatranscriptome and metaproteome analysis. In the lab, two of these

replicate bloom samples were frozen for proteome analysis. A third replicate sample from this field site was extracted for metatranscriptome analysis as described above.

Samples for protein analysis were thawed and then centrifuged for 30 min at 14,500 rpm 4 °C and stored on ice. The RNAlater supernatant was decanted from the pellet and protein material concentrated by 5000 MWCO filter (Sartorius Stedim Biotech Vivaspin) spun at RT, at 7,000 rpm and rinsed once with 2 mL 0.1 M Tris-HCl until 200 µL of liquid remained. Rinsed supernatant was then combined with the original pellet. Two mL of lysis buffer (1% SDS, 0.1 M Tris/HCl pH 7.5, 10 nM EDTA) was added and the slurry was transferred to a larger tube. Tubes were incubated at 95 °C for 10 min, and then incubated at RT for 1h at 350 rpm. The protein extract was then concentrated in a 5000 MWCO Vivaspin column to 300 µL. Samples were then precipitated in 1200 µL of 50:50 methanol:acetone with HCl (0.5 mM). Proteins were extracted, digested, and purified following the lab methods, and then identified on a Q-Exactive Orbitrap mass spectrometer using a Michrom Advance CaptiveSpray source using 1-dimension of chromatographic separation (1D) as described above for the culture studies.

An additional 2-dimensional chromatographic (2D) separation scheme was employed to provide deeper metaproteome results. In this case the samples were analyzed with a Thermo Fusion mass spectrometer following online 2-dimension active modulation liquid chromatography using a Dionex Ultimate3000 RSLCnano system with an additional RSLCnano pump. The first column separation utilized a nonlinear 8 hour pH = 10 gradient (10 mM Ammonium Formate and 10mM ammonium Formate in 90% acetonitrile) on a PLRP-S column (200 µm x 150 mm, 3 µm bead size, 300Å pore size, NanoLCMS Solutions). The eluent was diluted inline (10 µL/min 0.1% Formic acid) then trapped and eluted every 30 min on alternating dual traps (300 µm x 5 mm, 5 µm bead size, 100 Å pore size, C18 PepMap100, Thermo Scientific). The alternating traps were

eluted at 500 nL/min onto a C18 column (100 μm x 150 mm, 3 μm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH packed in a New Objective PicoFrit column) with a 30 min nonlinear gradient (0.1% Formic Acid and 0.1% Formic Acid in 99.9% Acetonitrile) on a Thermo Flex ion source attached to the mass spectrometer.

Identification of proteins within environmental samples is a challenging endeavor due to the need to map spectra to peptide sequences that are representative of the significant natural sequence diversity. To maximize our opportunities for protein identification we applied three multiple sequence databases to the peptide to spectrum matching analyses for the mass spectra collected on this field sample. These are listed in Supplementary Table 2 and include: 1) a combined version of the translated transcriptome database for both *Phaeocystis* strains derived from the culture laboratory studies described above (Database #1"; assembly_1374_and_1871_combined.orf.fasta), 2) a parallel metatranscriptome that was generated from the same Ross Sea field sample analyzed for metaproteomes (Database #2; this transcriptome is a combination of eukaryotic and prokaryotic communities derived from total RNA and poly(A) enriched RNA sequencing [2005_bloom_orfs_aa.fasta)]), and 3) five polar bacterial genomes assembled from a recent polar metagenome study in the Amundsen Sea downloaded from NCBI and concatenated as polar bacteria database, including the bacterium *Flavobacteriaceae bacterium* ASP10-09a, *Cryomorphaceae bacterium* ASP10-05a, *Marine Gamma Proteobacterium* ASP10-03a, *Rhodobacteraceae bacterium* ASP10-04a, *Oceanospirillaceae bacterium* ASP10-02a (Database #3; polar_genomes_110515.fasta) (Delmont et al., 2014).

This application of multiple databases allowed for an increased number of protein identifications and has some specific advantages. For example, use of the transcriptomes from culture studies (Database #1 above) allows direct comparison to the culture experiments. Database

#2 and #3 allow an analysis of a broader diversity of the eukaryotic and prokaryotic components than can be captured by laboratory isolates. This was particularly evident when searching using a compiled microbial metagenomics database from temperate and tropical regions that produced few results of highly conserved proteins (indicative of commonly shared peptides). Finally, use of Database #2 allows a paired metatranscriptome-metaproteome for a taxonomically more representative dataset.

In this analysis, we calculated the number of unique tryptic peptides identified from applying each database as a basal measure of metaproteome diversity due to the potential for tryptic peptides to be shared between taxa, a common challenge for protein inference in metaproteomic samples that contain many organisms. Although the occurrence of identical peptides between diverse taxonomic group is relatively rare in marine prokaryotes, typically less than a few percent of total tryptic peptides (Denef et al., 2007; Saito et al., 2015), it can result in ambiguity when counting at the protein level. Moreover, redundancies of peptide sequences can occur in the metatranscriptomic reference databases and this is resolved also by focusing on unique peptides. While our protein and peptide identifications are based on rather conservative statistical thresholds, future improvements in mass spectrometry informatics specialized in metaproteomic approaches should expand this depth of protein discovery considerably.

Analyses using the culture transcriptomes to strains 1871 and 1374 (Database #1) were conducted using the Proteome Discoverer (SEQUEST) and Scaffold to settings of protein and peptide thresholds of 99% and 99% with 1 minimum peptide and a parent mass tolerance of 10 ppm, resulting in a decoy peptide false discovery rate (FDR) of 0.6 and 0.17% for 2D and 1D analyses respectively 1.46%. Analyses using the paired metatranscriptome-metaproteome (Database #2) were conducted using the Proteome Discoverer (SEQUEST) and Scaffold to settings

of protein and peptide thresholds of 99% and 99% with 1 minimum peptide and a parent mass tolerance of 10ppm, resulting in a 0.1% and 0.70% decoy FDR for 2D and 1D analyes. Analyses of the bacterial metaproteome (Database #3) were conducted using the double search metaproteomic analysis method (Jagtap et al., 2013) where proteins sequences identified in a first pass by peptide mapping algorithm SEQUEST with protein and peptide probability settings of 80% were exported and used as a smaller database on a second pass with more stringent settings of Delta CN score of 0.1, and XCorr values of 1.8, 2.5, 3.5, 3.5 for +1 - +4 precursors, respectively, and a parent mass tolerance of 10 ppm (Supplementary Data 2).

The Ross Sea microbial ecosystem represents a complex coastal ecosystem. Nevertheless targeted metaproteomics appears to have sufficient resolution to identify and quantitate numerous proteins of interest. To demonstrate this capability, a spectral database of the proteins of interest was created in Scaffold from Proteome Discoverer 1.4 matched spectra, and MS1 filtering was performed with Skyline (Schilling et al., 2012) (Skyline (64-bit) 3.5.0.9319, MacCoss Lab, Department of Genome Sciences, UW). Successful targeting of peptide corresponding to proteins of biological and biogeochemical interest (uncalibrated) was demonstrated in Supplementary Figures 4 through 10 and their targeting parameters are presented in Supplementary Table 3.

Field pigment analyses

Seawater samples were filtered onto glass fiber filters (Whatman GF/F) and stored in liquid nitrogen until analysis. Samples were analyzed on an Agilent 1100 HPLC (High Performance Liquid Chromatography) system with diode array and fluorescence detection. Elution gradient and protocols were described in detail previously (DiTullio et al., 2003).

Data availability

Phaeocystis antarctica RNA sequence data reported in this paper have been deposited in the NCBI sequence read archive under BioProject accession no. PRJNA339150, BioSample accession nos. SAMN05580299 – SAMN05580303. Ross Sea metatranscriptomes have been deposited under BioProject accession no. PRJNA339151, BioSample accession nos. SAMN05580312 – SAMN05580313. Proteomic data from the lab and field components was submitted to the Pride database (Project Name: *Phaeocystis antarctica* CCMP 1871 and CCMP 1374, Ross Sea Phaeocystis bloom, LC-MSMS; Project accession: PXD005341; Project DOI: 10.6019/PXD005341).

Supplementary Table 1. Response of known iron-related proteins in *Phaeocystis antarctica* cultures.

Protein	Strain	Contig ID	Relationship to iron	Figure
Plastocyanin	1871	contig_73520_249_1003_-	Inverse	4
Flavodoxin	1871	contig_31444_1_606_+	Inverse	S1
ISIP3	1871	contig_85414_1_437_+	Inverse	S1
Ferredoxin	1871	contig_106114_64_1011_+	Positive	S1
ISIP2A	1871	contig_112549_1_768_-	Positive	S1
Ferredoxin	1871	contig_119121_30_1093_+	Positive	4
ISIP2A	1871	contig_14972_951_2648_-	Positive	S1
ISIP2A	1871	contig_18112_1_532_-	Positive	S1
ISIP2A	1871	contig_25911_1_794_+	Positive	4
ISIP2A	1871	contig_30604_1_905_-	Positive	S1
ISIP2A	1871	contig_67880_1_1075_+	Positive	S1
ISIP3	1871	contig_68024_264_2093_-	No Trend	S1
ISIP3	1374	contig_81580_292_2611_-	Inverse	4
ISIP2A	1374	contig_179615_981_2023_-	Inverse	S1
ISIP2A	1374	contig_16665_1_726_-	Inverse	S1
Flavodoxin	1374	contig_202625_47_661_+	No Trend	S1
Ferredoxin	1374	contig_109246_59_1006_+	Inverse	4
ISIP3	1374	contig_81593_985_1876_-	Inverse	4
ISIP3	1374	contig_81593_209_919_-	Inverse	4
Ferredoxin	1374	contig_144194_58_1008_+	No Trend	S1
Flavodoxin	1374	contig_63273_96_794_+	Inverse	4
ISIP2A	1374	contig_180064_1_1085_+	Positive	4
ISIP3	1374	contig_32359_1_1574_-	Positive	4
ISIP2A	1374	contig_84443_1_798_-	Positive	4
ISIP2A	1374	contig_84910_1_781_+	Positive	S1

Supplementary Table 2. Databases used for metaproteome analysis of *Phaeocystis* Ross Sea 2005 bloom samples.

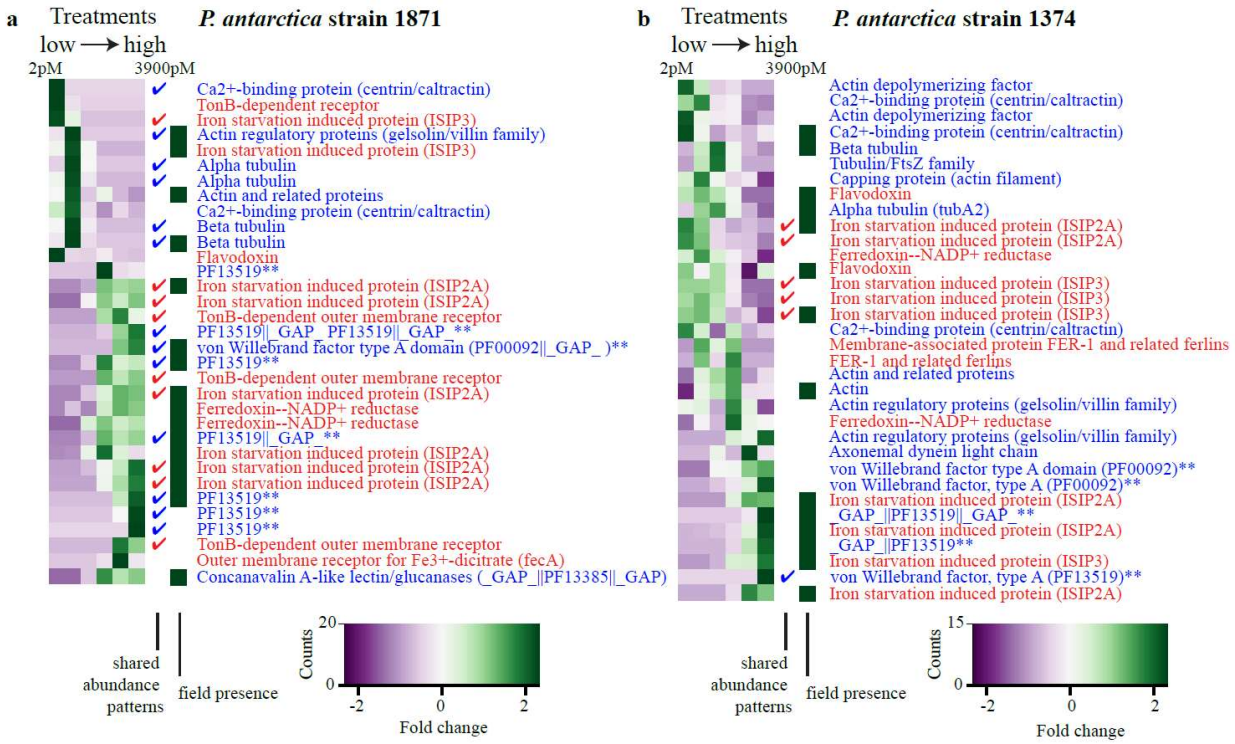
Database #	Description	Database size (Protein entries)	Total Identifications 1D Peptide (Protein)	Total Identifications 2D Peptide (Protein)
1	Transcriptomes from *Phaeocystis* cultures strains 1871 and 1374	191496	2103 (912)	3816 (1545)
2	Parallel metatranscriptome from 2005 December 30th Ross Sea Sample (combined Total RNA and Poly(A) enrichment)	208126	1520 (859)	3210 (1474)
3	Antarctic bacterial metagenomics from Amundsen Sea (Delmont et al., 2014)	12824	186 (92)	237 (102)

Supplementary Table 3. Identified putative protein biomarkers in the Ross Sea metaproteome used for example targeted demonstrations in Supplementary Fig. 4 – 10.

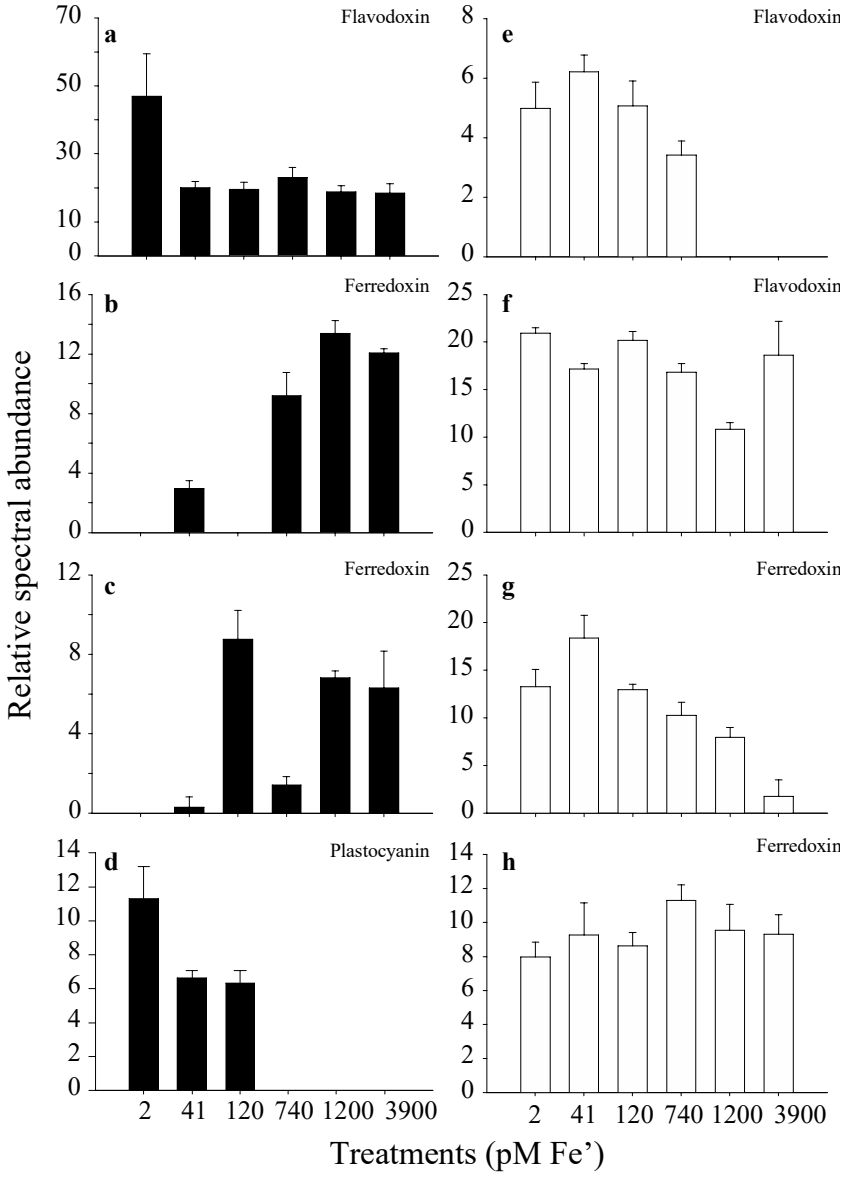
Figure	Protein	ID (contig)	Peptide sequence	mass	charge	db 1	db 2	db 3
10	flavodoxin	31444_1_606_+	SGTAWDEFIYGDLAGLDLK	2071.0019	2	x	x	
10	"	31444_1_606_+	AWIAQIK	829.4903	3	x	x	
S4	flavodoxin	119121_30_1093_+	VGEYADEVFTR	1285.6059	2		x	
S4	"	119121_30_1093_+	GICSNFLCDSKPGDEVK	1925.8723	3		x	
S5	plastocyanin	155134_121_481_-	GGPHNVVFDEDAIPAGVSQEK	2166.0462	3		x	
S5	"	114328_8_706_-	GDSITWINNK	1147.5742	2		x	
S5	"	114328_8_706_-	GGPHNVVFVEDAIPK	1578.8275	3	x	x	
S6	ISIP2A	175280_1_1007_+	VLNLFQVGQNAIR	1472.8453	2		x	
S7	sodA	73278_189_657_-	IDAAFGSYDDFK	1349.6129	2	x	x	
S8	bacterioferritin	483_71_526_+	IPIVSWLK	955.5975	2		x	
S8	"	13894_187_684_+	LAAEENNVLIEEYAR	1733.8705	2		x	
S9	TonB	197536_1866_2873_-	DGASAVYGSDAVAGVVNIITR	2036.0527	2		x	
S9	"	197536_1866_2873_-	GPEFLDSWR	1107.5338	2		x	

* where x's within columns db 1-3 refer to the presence of the peptide within *Phaeocystis* metatranscriptome, Ross Sea metatranscriptome, and Amundsen Sea metagenome databases, respectively (see Supplementary Table 2).

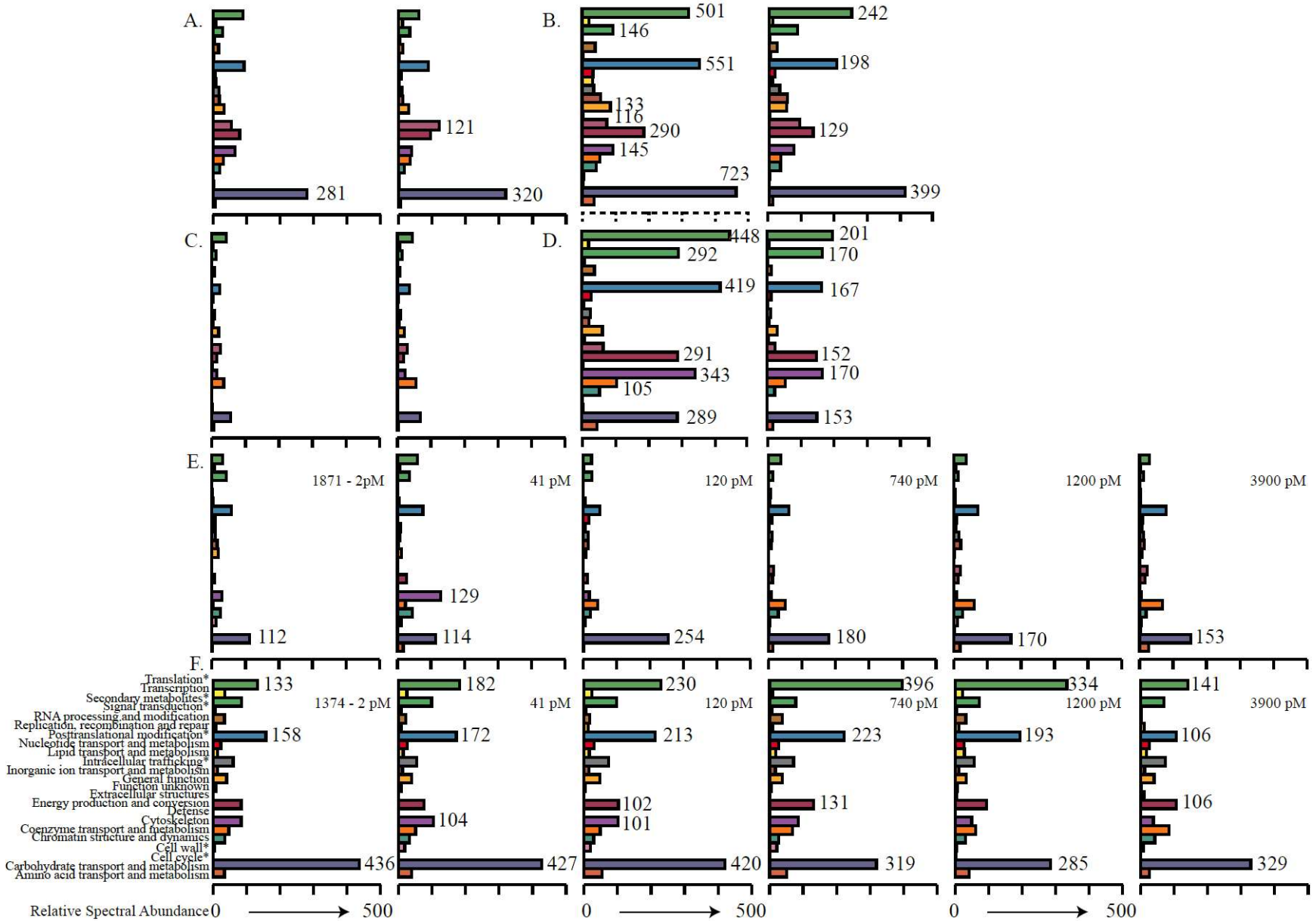
Supplementary Fig. 1. Heatmaps for additional hand-selected iron (red) and structurally related (blue) proteins for *P. antarctica* strain 1871 (a) and strain 1374 (b). Heatmaps represent the relative protein abundance for the six treatments for each strain; the darker green color indicates a greater relative abundance compared to the purple treatments. The "shared abundance patterns" column features a check-mark when a shared response to changes in iron availability between the relative protein abundance and the transcript abundance was observed (for example, both transcripts and proteins have a higher abundance under high iron compared to low iron growth [or] both transcripts and proteins have a higher abundance under low iron compared to high iron growth). The "field presence" column indicates whether or not that protein was detected in the field metaproteome (annotated using Database #1 and including 1D and 2D analysis). Protein annotations are based on KEGG, KOG, and Pfam descriptions. Proteins with putative glycoprotein, adhesion, and/or ligand-binding domains are marked with a double asterisk.



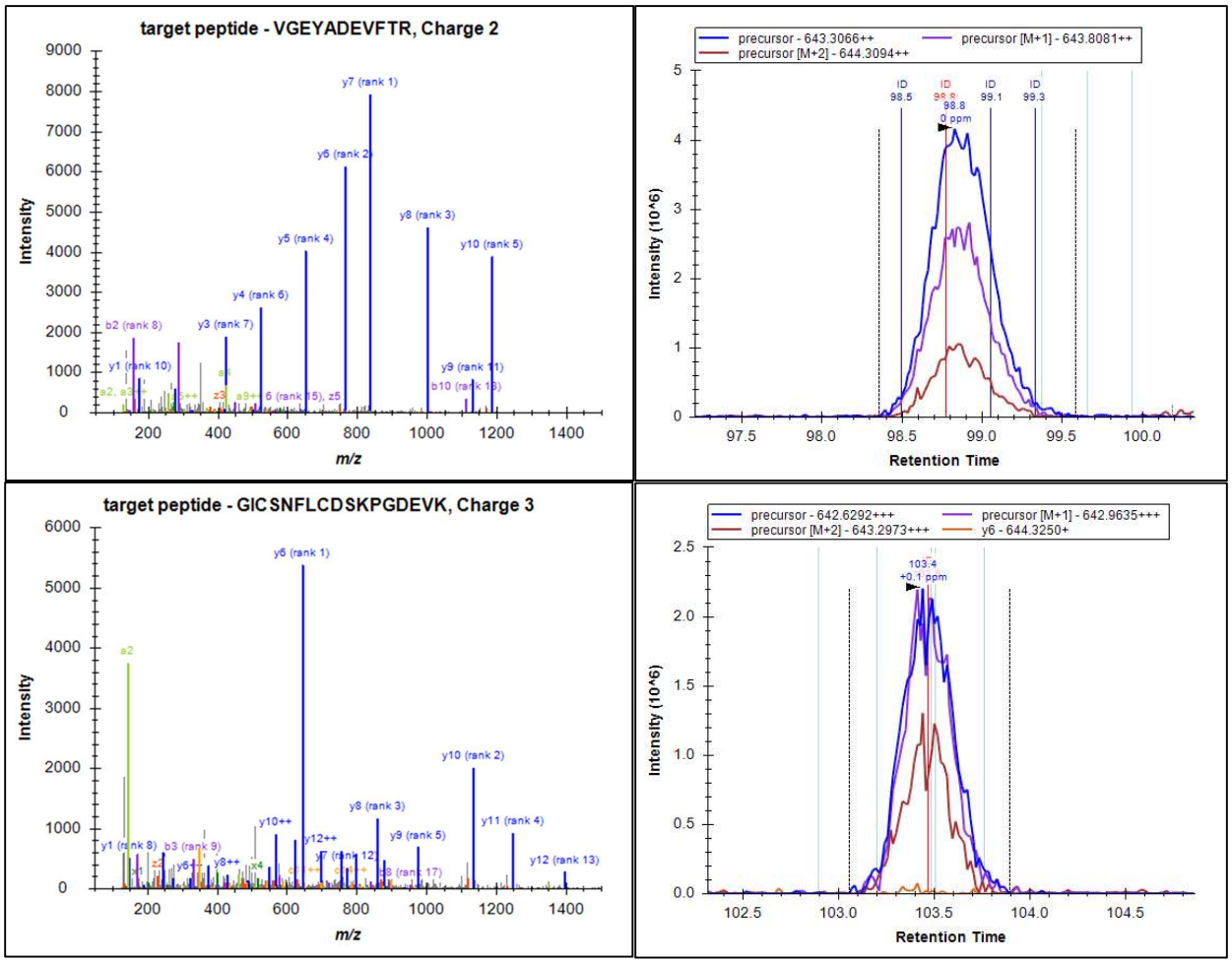
Supplementary Fig. 2. Bar graphs for putative flavodoxin, ferredoxin, and plastocyanin proteins identified from the lab experiments. (*a-d*) Black bars, strain 1871; (*e-h*) White bars, strain 1374. Error bars represent the standard deviation for three technical replicates (n=3).



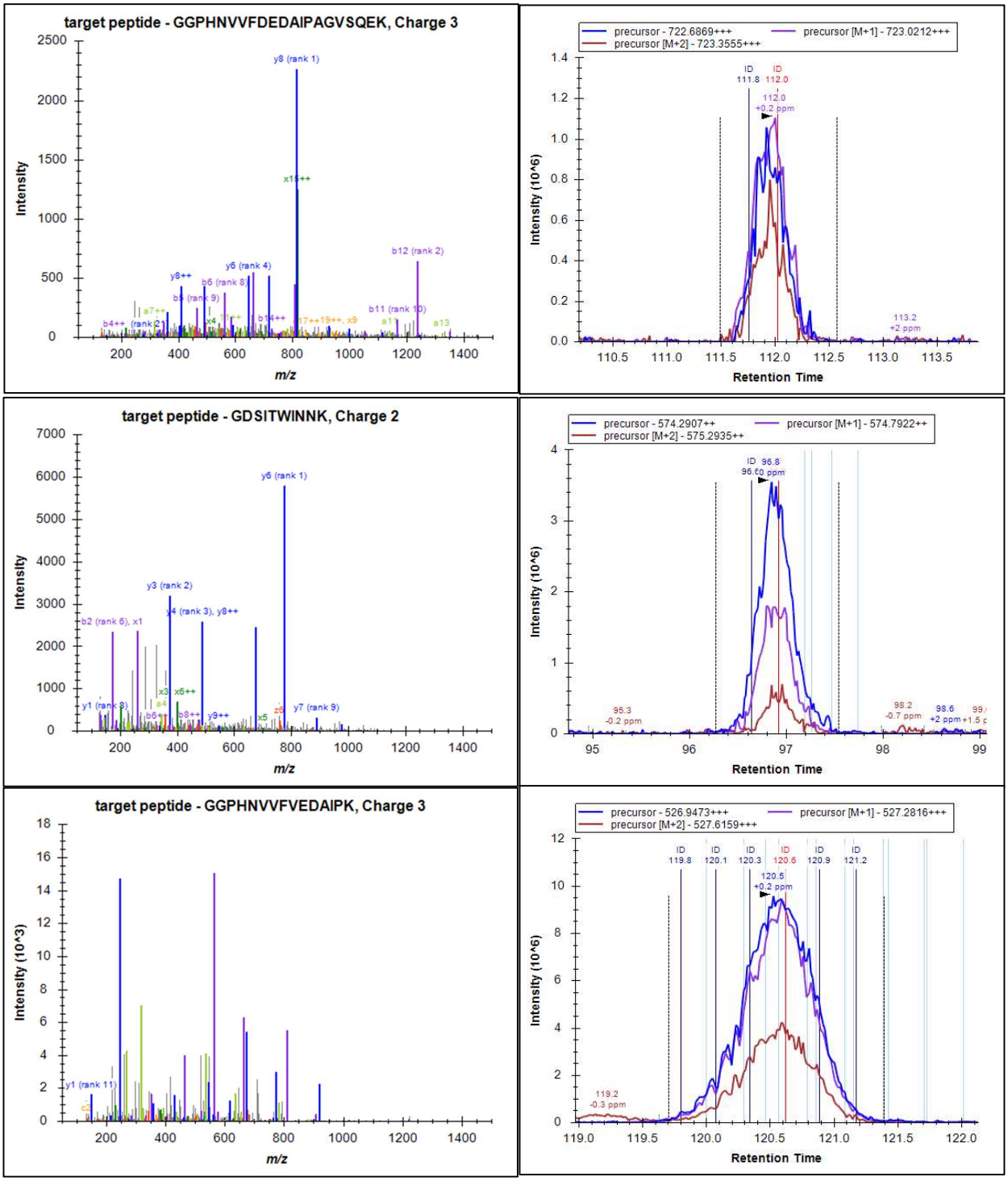
Supplementary Fig. 3. KOG class annotations (y-axis) for all (meta)proteomes covered in this study. Field metaproteomes collected from the Ross (*a*: left and right panels are biological replicates, annotated using the 1D-Database #1; *b*: left and right panels are biological replicates annotated using the 2D-Database #1; *c.* left and right panels are biological replicates annotated using the 1D-Database #2; *d:* left and right panels are biological replicates annotated using the 2D-Database #2). Laboratory proteomes for *P. antarctica* strain 1871 (*c*), and strain 1374 (*d*) across the six laboratory treatments (left panel, 2 pM Fe' to right panel, 3900 pM Fe'). Total number of spectral counts for each annotation (≥ 100) are shown next to each bar. X-axis, "relative spectral abundance." The dashed axis (*b*, left panel) represents a scale from 0 to 800; all other axes in figure are 0 to 500.


A.
B.
C.
D.
E.
F.
281
320
501
146
551
133
116
290
145
723
242
198
129
399
448
292
419
291
343
105
289
201
170
167
152
170
153
1871 - 2pM
41 pM
120 pM
740 pM
1200 pM
3900 pM
112
129
114
254
180
170
153
1374 - 2 pM
133
158
436
182
172
104
427
230
213
102
101
420
396
223
131
319
334
193
285
141
106
106
329
Translation*
Transcription
Secondary metabolites*
Signal transduction*
RNA processing and modification
Replication, recombination and repair
Posttranslational modification*
Nucleotide transport and metabolism
Lipid transport and metabolism
Intracellular trafficking*
Inorganic ion transport and metabolism
General function
Function unknown
Extracellular structures
Energy production and conversion
Defense
Cytoskeleton
Coenzyme transport and metabolism
Chromatin structure and dynamics
Cell wall*
Cell cycle*
Carbohydrate transport and metabolism
Amino acid transport and metabolism
Relative Spectral Abundance
0
500
0
500
0
500

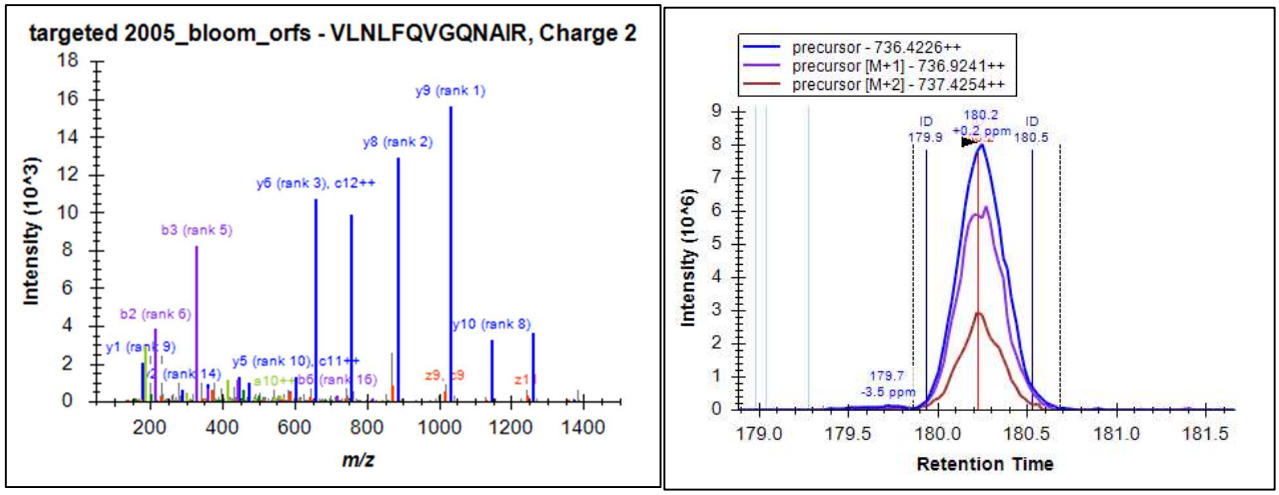
Supplementary Fig. 4. Example spectra and chromatograms of fragment ions for flavodoxin identified from the Ross Sea metaproteome sample (peptides found within Database #1, 1871, contig_119121_30_1093_+).



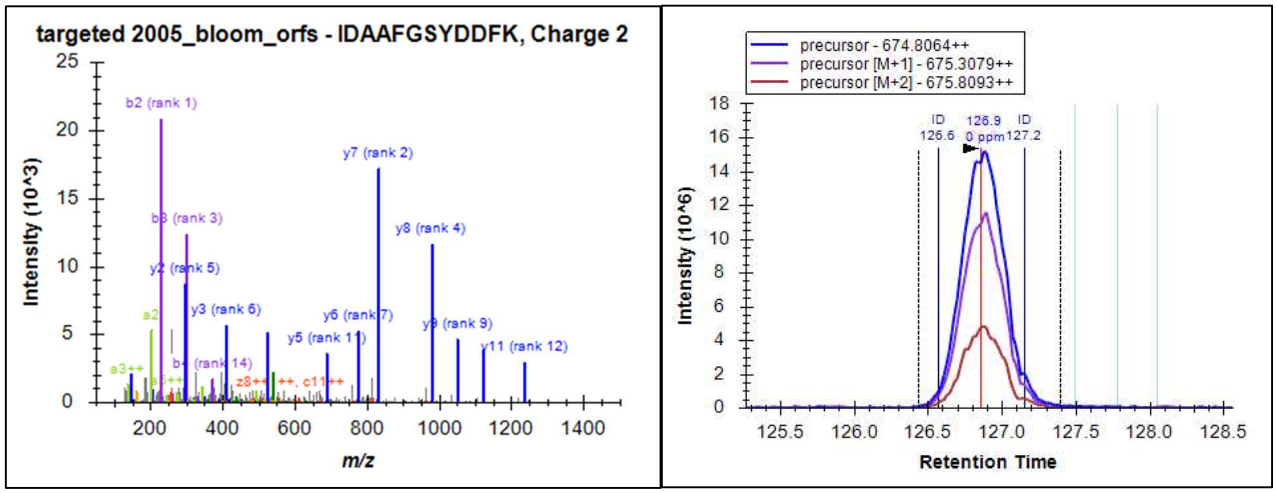
Supplementary Fig. 5. Example spectra and chromatograms of fragment ions for plastocyanin identified from the Ross Sea metaproteome sample (peptides found within Database #2, contig_155134_121_481_- [top], contig_114382_8_706_- [middle and bottom]).



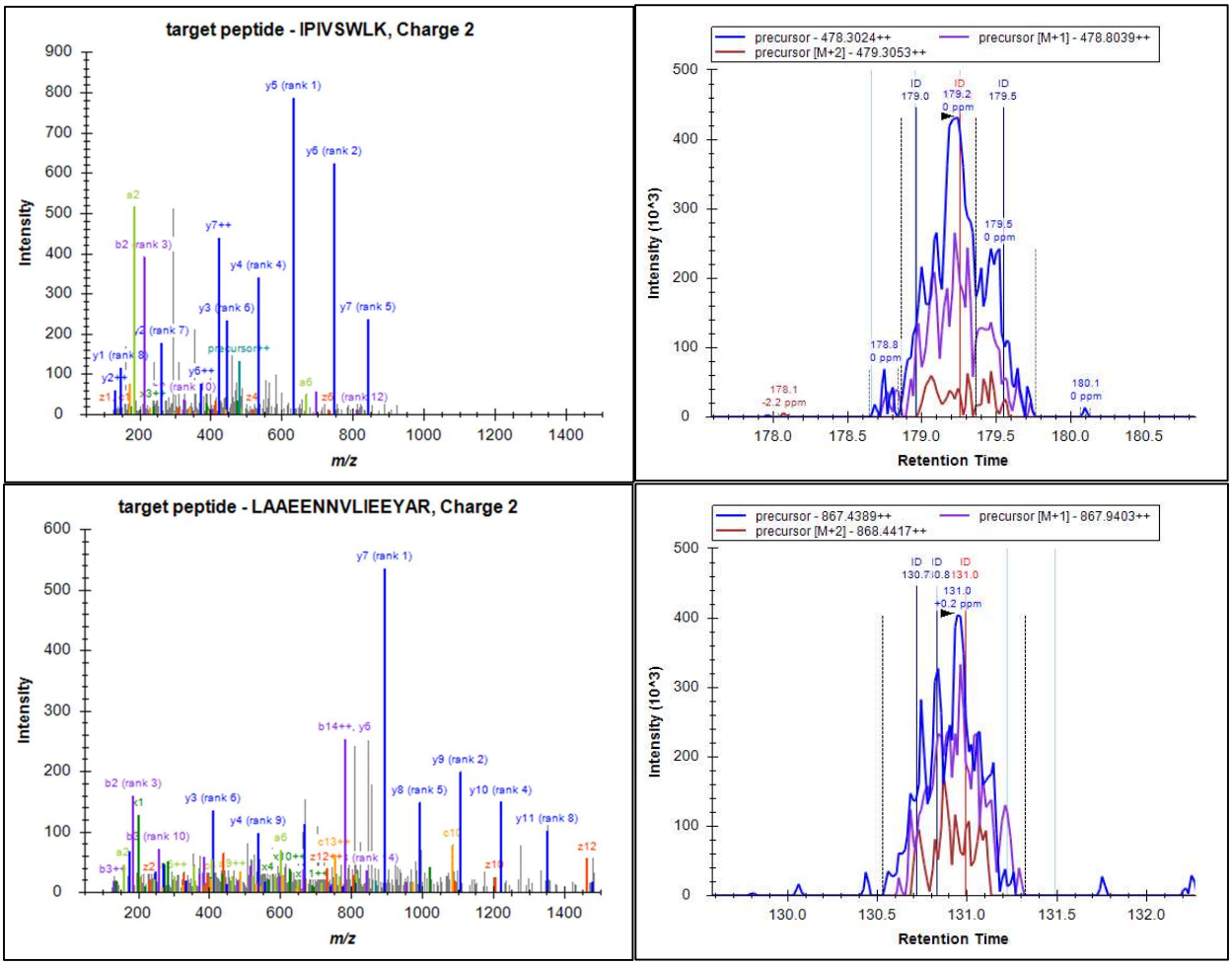
Supplementary Fig. 6. Example spectra and chromatograms of fragment ions for ISIP2A iron starvation induced protein identified from the Ross Sea metaproteome sample (peptide found within Database #2, contig_175280_1_1007_+).



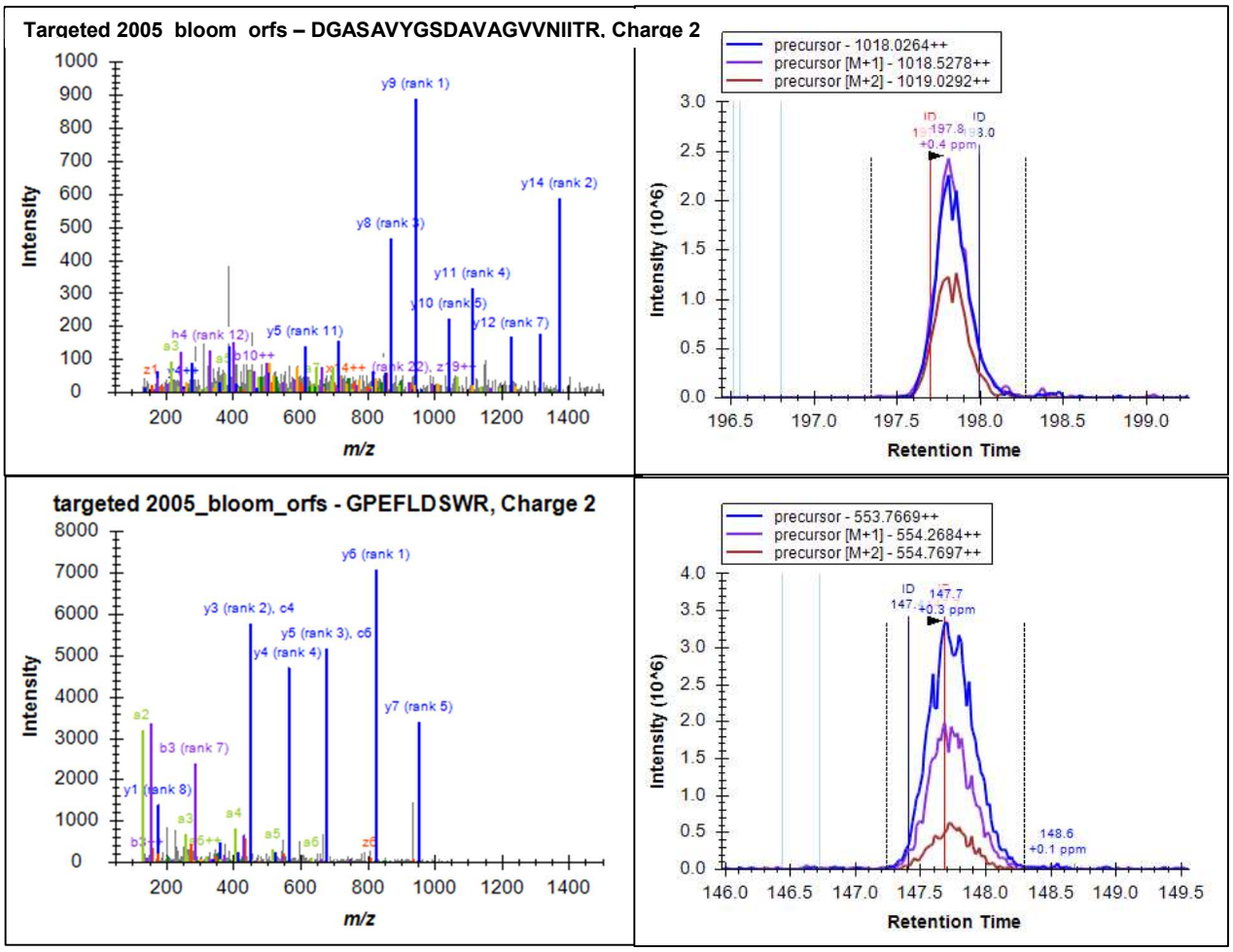
Supplementary Fig. 7. Example spectra and chromatograms of fragment ions for sodA; superoxide dismutase, manganese identified from the Ross Sea metaproteome sample (peptide found within Database #1, 1374, contig_73278_189_657_-).



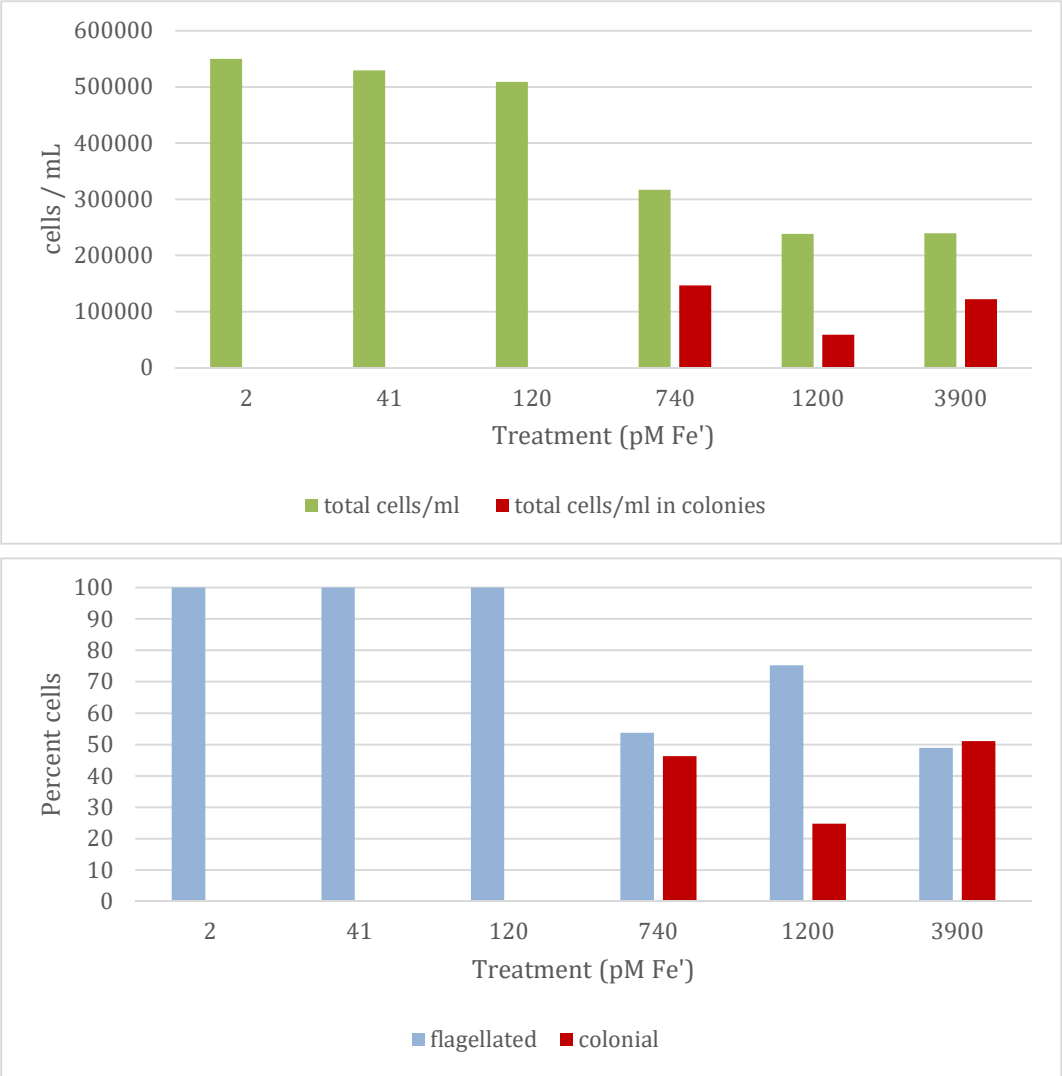
Supplementary Fig. 8. Example spectra and chromatograms of fragment ions for bacterioferritin from the Ross Sea metaproteome sample (peptides found within Database #2, contig_483_71_526_+ [top], contig_13894_187_684_+ [bottom] – same as Database #3, WP_046485702).



Supplementary Fig. 9. Example spectra and chromatograms of fragment ions for TonB-dependent receptor from the Ross Sea metaproteome sample (peptides found within Database #2, contig_197536_1866_2873).



Supplementary Fig. 10. An additional experiment on *Phaeocystis antarctica* strain 1871 across six iron treatments using the same approach and methods as that shown in Figure 3 showing the same trend in colony formation at higher iron concentrations. In this experiment, cells were counted at the time of harvest. Total cells and colony cells were counted and flagellated cells calculated by difference. Note that the number of cells is higher in the lowest three iron treatments due to being given a longer growth period prior to harvest, and since biomass is greater in colonial cells due to their larger cell size (see Figure 3e).



Supplementary References

Bertrand, E. M., McCrow, J. P., Moustafa, A., Zheng, H., McQuaid, J. B., Delmont, T. O., Post, A. F., Sipler, R. E., Spackeen, J. L., Xu, K., Bronk, D. A., Hutchins, D. A. and Allen, A. E.: Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge, Proceedings of the National Academy of Sciences, 112(32), 9938–9943, doi:10.1073/pnas.1501615112, 2015.

Delmont, T. O., Eren, A. M., Vineis, J. H. and Post, A. F.: Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica, Frontiers in Microbiology, 6, doi:10.3389/fmicb.2015.01090, 2015.

Delmont, T. O., Hammar, K. M., Ducklow, H. W., Yager, P. L. and Post, A. F.: *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya, Frontiers in Microbiology, 5, 646, 2014.

Denef, V. J., Shah, M. B., VerBerkmoes, N. C., Hettich, R. L. and Banfield, J. F.: Implications of strain- and species-level sequence divergence for community and isolate shotgun proteomic analysis, J. Proteome Res., 6(8), 3152–3161, doi:10.1021/pr0701005, 2007.

DiTullio, G. R., Geesey, M. E., Jones, D. R., Daly, K. L., Campbell, L. and Smith, W. O.: Phytoplankton assemblage structure and primary productivity along 170º W in the South Pacific Ocean, Marine Ecology Progress Series, 255, 55–80, 2003.

Jagtap, P., Goslinga, J., Kooren, J. A., McGowan, T., Wroblewski, M. S., Seymour, S. L. and Griffin, T. J.: A two-step database search method improves sensitivity in peptide sequence matches for metaproteomics and proteogenomics studies, Proteomics, 13(8), 1352–1357, doi:10.1002/pmic.201200352, 2013.

Lê, S., Josse, J. and Husson, F.: FactoMineR: an R package for multivariate analysis, Journal of Statistical Software, 25(1), 2008.

Li, H.: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv, 2013.

Lu, X. and Zhu, H.: Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins, Molecular & Cellular Proteomics, 4(12), 1948–1958, doi:10.1074/mcp.M500138-MCP200, 2005.

Noble, A. E., Lamborg, C. H., Ohnemus, D. C., Lam, P. J., Goepfert, T. J., Measures, C. I., Frame, C. H., Casciotti, K. L., R, D. G., Jennings, J. and Saito, M. A.: Basin-scale inputs of cobalt, iron, and manganese from the Benguela-Angola front to the South Atlantic Ocean, Limnology and Oceanography, 57(4), 989–1010, doi:10.4319/lo.2012.57.4.0989, 2012.

Podell, S. and Gaasterland, T.: DarkHorse: a method for genome-wide prediction of horizontal gene transfer, Genome Biology, 8, R16, doi:10.1186/gb-2007-8-2-r16, 2007.

Price, N. M., Harrison, G. I., Hering, J. G., Hudson, R. J., Nirel, P. M. V., Palenik, B. and Morel, F. M. M.: Preparation and chemistry of the artificial algal culture medium aquil, Biological Oceanography, 6, 5–6, 1989.

Rho, M., Tang, H. and Ye, Y.: FragGeneScan: predicting genes in short and error-prone reads, Nucleic Acids Research, 38(20), e191–e191, doi:10.1093/nar/gkq747, 2010.

Saito, M. A., Dorsk, A., Post, A. F., McIlvin, M. R., Rappé, M. S., R, D. G. and Moran, D. M.: Needles in the blue sea: Sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome, Proteomics, 15(20), 3521–3531, doi:10.1002/pmic.201400630, 2015.

Saito, M. A., Goepfert, T. J., Noble, A. E., Bertrand, E. M., Sedwick, P. N. and DiTullio, G. R.: A seasonal study of dissolved cobalt in the Ross Sea, Antarctica: micronutrient behavior, absence of scavenging, and relationships with Zn, Cd, and P, Biogeosciences, 7, 4059–4082, doi:10.5194/bg-7-4059-2010, 2010.

Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M., Frewen, B. E., Cusack, M. P., Sorensen, D. J., Bereman, M. S., Jing, E., Wu, C. C., Verdin, E., Kahn, C. R., MacCoss, M. J. and Gibson, B. W.: Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline, Molecular & Cellular Proteomics, 11, 202–214, 2012.

Schmieder, R., Lim, Y. W. and Edwards, R.: Identification and removal of ribosomal RNA sequences from metatranscriptomes, Bioinformatics, 28(3), 433–435, 2012.

Sedwick, P. N., Marsay, C. M., Sohst, B. M., Aguilar Islas, A. M., Lohan, M. C., Long, M. C., Arrigo, K. R., Dunbar, R. B., Saito, M. A., Smith, W. O. and DiTullio, G. R.: Early season depletion of dissolved iron in the Ross Sea polynya: Implications for iron dynamics on the Antarctic continental shelf, Journal of Geophysical Research, 116(C12019), 2011.

Sunda, W. and Huntsman, S.: Effect of pH, light, and temperature on Fe–EDTA chelation and Fe hydrolysis in seawater, Marine Chemistry, 84(1-2), 35–47, doi:10.1016/S0304-4203(03)00101-4, 2003.

Sunda, W. G. and Huntsman, S. A.: Iron uptake and growth limitation in oceanic and coastal phytoplankton, Marine Chemistry, 50, 189–206, 1995.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S. and Schwartz, M.: gplots: Various R programming tools for plotting data. 2009.

Wu, Z., Jenkins, B. D., Rynearson, T. A., Dyhrman, S. T., Saito, M. A., Mercier, M. and Whitney, L. P.: Empirical bayes analysis of sequencing-based transcriptional profiling without replicates, BMC Bioinformatics, 11, 564, doi:10.1186/1471-2105-11-564, 2010.