Journal of
Emerging Investigators

Phytoplankton Plastid Proteomics: Cracking open Diatoms to Understand Plastid Biochemistry under Iron Limitation

Skyler J. Nunn[1,2], P. Dreux Chappell[3], Kristofer Gomes[4], Anasthasia Bonderenko[2,5], Bethany D. Jenkins[4], Brook L. Nunn[2]*

[1]Telluride High School; 725 West Colorado Avenue, Telluride CO 81435. [2]University of Washington, Department of Genome Sciences; Seattle WA 98195. [3]Old Dominion University, Ocean, Earth & Atmospheric Sciences; Norfolk, VA 23529. [4]University of Rhode Island, Department of Cell and Molecular Biology, Rhode Island 02881. [5]Newport High School, Newport, WA 99156. *Corresponding author, brookh@uw.edu

Summary

Diatoms, such as *Thalassiosira pseudonana*, are important oceanic primary producers, as they sequester carbon dioxide (CO_2) out of the atmosphere, die, and precipitate to the ocean floor. In many areas of the world's oceans, phytoplankton, such as diatoms, are limited in growth by the availability of iron (Fe). Fe is an essential nutrient for phytoplankton, as it is central in the electron transport chain component of photosynthesis. Through this study, we examined if Fe-limitation makes a significant difference in the proteins expressed within the chloroplast, the power source for diatoms. Here, we utilized a new plastid isolation technique specific to diatoms and completed 14 mass spectrometry experiments to determine how many proteins transit the plastid membrane, if there are any differences in the expressed proteomes within the plastid grown under Fe-replete and Fe-limited conditions, and what those differences are. Over 900 unique proteins were identified from the isolated plastids, and cluster analyses of the data verified that statistical differences are present between the Fe-replete and Fe-limited growth conditions. Furthermore, our plastid proteome is in agreement with many of the recognized proteins previously discovered in land plant plastids, suggesting the isolation method followed by proteomic mass spectrometry is valid and sensitive. Through the isolation and analysis of plastid proteins, as shown here, scientists can now better identify which nutrients and/or trace metals directly affect diatom photosynthetic capacity so they can design better experiments to increase CO_2 fixation rates.

Received: April 25, 2016; **Accepted:** December 11, 2016; **Published:** February 10, 2017

Introduction

The global epidemic of increased carbon dioxide (CO_2) in the atmosphere has brought many researchers together in an attempt to find a solution. One such idea for sequestering CO_2 out of the atmosphere is to employ phytoplankton, a dominant oceanic primary producer. Phytoplankton are limited by Fe in vast areas of the world's oceans, and it has been suggested that the anthropogenic act of iron (Fe) fertilization could increase bloom size and the resulting drawdown of CO_2 [1-5]. The act of manually enriching the oceans with Fe is contentious, as the direct and long-term effects Fe additions will have on the marine ecosystems is unknown [6, 7]. Diatoms constitute 40% of marine organic carbon and account for as much as 30% of global carbon fixation [8]. Fe fertilization is well documented to increase diatom bloom size and the amount of CO_2 removed from the atmosphere, yet the direct physiological effect Fe has on diatoms is still being explored. In these areas of limited Fe, diatoms have diminished photosynthetic abilities and reduced chlorophyll contents. Fe plays a primary role in many proteins involved in energy production and utilization [9, 10]. Many of these essential functions are predicted to take place in the mitochondria and chloroplasts, yet little is known about the distributions of these functions in diatoms [11-14]. Although the plastid is the location within the cell that generates cellular energy via photosynthesis, to date it is unknown if all the proteins within the plastid are constitutively expressed or if the plastid needs to modify the proteome to adapt to environmental stresses, such as Fe-limitation.

Based on previous research by Nunn *et al.* [10], we hypothesized that diatoms express different proteins in order to adapt to environmental stresses and that this expression will be observable using mass spectrometry techniques. To understand how chloroplast protein function may be impacted by low Fe, we grew cultures of diatoms in Fe-replete or Fe-limited seawater. The cells were harvested for chloroplast isolation at mid-exponential growth. Because diatoms are enclosed in a silica shell known as a frustule, it is difficult to disrupt the cells and maintain an intact chloroplast. Traditional means to disrupt cells typically use a high frequency sonicating probe that shatters the shell and

all internal organellar components. We used a recently developed protocol that utilizes a chemical-based method to dissolve the outer silica shell, leaving the internal protoplast membrane and organelles unharmed. Organelles were then isolated using Percoll density gradients and isolated fractions were digested with trypsin and analyzed using mass spectrometry-based proteomics. Recent bioinformatical approaches have examined plastid localization [14], but to our knowledge, this is the first experimental analysis of its kind on diatoms. Since there are only 127 proteins encoded in the *T. pseudonana* plastid genome [12], we hypothesized that the most highly expressed proteins within the plastid are from the plastid-encoded proteome, with hundreds of additional proteins translated from the diatom's mitochondrial- or nuclear-encoded genome within the plastid to ensure functionality. The second objective of the research was to determine if proteins transcribed from either the mitochondrial or nuclear genome transit across the plastid membrane in order for the cell to photosynthesize and grow. Here, we present proteomic data from isolated chloroplasts to reveal the distribution of diatom plastid-localized processes under the stress of low Fe conditions. From the statistical analyses presented here, we determined that the distribution of proteins present within the plastids of diatom cells grown under Fe-replete and Fe-limited conditions exhibit significant differences. Further, we reveal which proteins drive the differences in the protein expression observed.

Results

Using mass spectrometry on isolated plastids from Fe-replete and Fe-limited cells, we determined that the cohort of proteins expressed in the two growth conditions exhibited significant differences. This discovery directly supports our hypothesis that diatoms express different proteins when faced with an environmental stress. Specifically, this was statistically revealed when the matrix of proteomic data and their relative quantities were examined using non-metric multidimensional scaling (NMDS; **Figure 1**). The NMDS analyses projects high-dimensional data (*i.e.*, abundances of 929 proteins) onto a new set of axes (ordinations), where the x-axis explains the greatest variation in the original data and the y-axis explains the second most variation. Further analysis of proteins predominantly contributing to the observed separation on NMDS axis 1 revealed that a plastid-localized GroEL chaperone (A0T0X0) is the primary component driving the differences observed between Fe-replete and Fe-limited diatom plastid proteomes on the NMDS plot. GroEL proteins are chaperonins involved in protein folding. Other proteins that were significantly more abundant in the Fe-replete samples and contributed to the separation on the NMDS plot included mitochondrial glycine decarboxylase P-protein (GdcP; B8BX31), aconitase (AcnB, B83CV9), pyruvate dehydrogenase (PdhB1, B8LC08), and 7 uncharacterized proteins (B8BPW0, B8BW88, B8C108, B8C4T7, B8CBZ7, B8CFJ8, B8CG91).

Although the *T. pseudonana* plastid-encoded genome has only 127 predicted proteins, our MS-based proteomic experiments identified 808 unique proteins in the Fe-limited cultures and 717 in the Fe-replete cultures (**Figure 2**). This is evidence that many of the proteins within the plastid have transited from the mitochondria, cytoplasm, or nucleus to the interior of the plastid; this finding supports our original hypothesis. Of the total 929 proteins, 212 proteins were uniquely identified in the Fe-

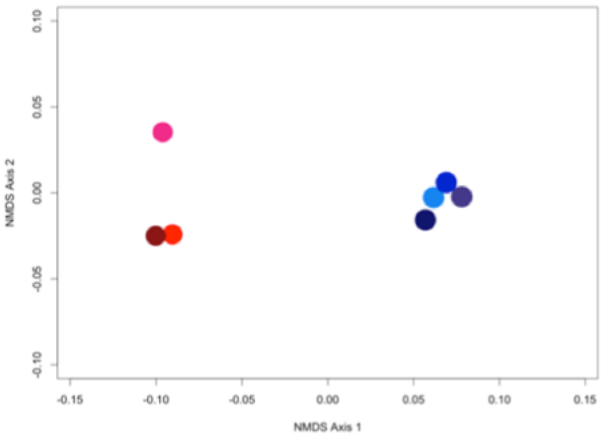

Figure 1. NMDS plot of protein expression data for biological splits and replicates of Fe-replete (shades of red) and Fe-limited (shades of blue) isolated plastids. Each point includes spectral count data averaged across two technical MS replicate analyses.

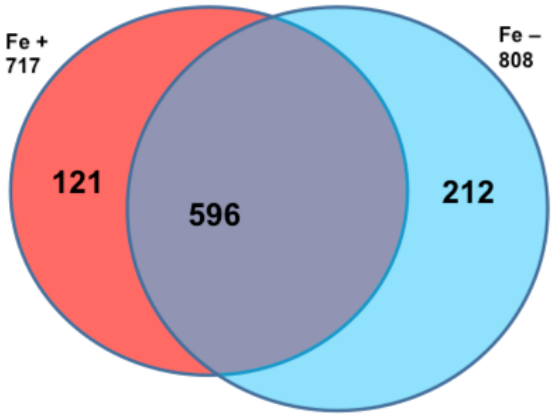

Figure 2. Venn Diagram illustrating overlap of proteins identified in plastids isolated from Fe-replete (Fe+) and Fe-limited (Fe-) cultures. Proteins were considered present within the sample condition (*i.e.*, Fe+ or Fe-) if each of the biological replicates or splits had > 2 spectral counts. Reported numbers within the diagram include all proteins identified from combined biological and technical replicates.

Biological Process GO Category	Count	%	*p*-value	UNIPROT Protein IDs
Generation of precursor metabolites and energy	9	8.65%	0.00824	B8BXS1, B5YN92, Q3S276, B8C0L7, B8C631, B8BR92, B8CAK5, B8C3V9, B8C0G1
Cellular macromolecular complex assembly	5	4.81%	0.01773	B5YMW7, B8CEQ7, B8BRR6, B8BS24
Glycolysis	4	3.85%	0.01827	B8BXS1, B5YN92, B8C0L7, B8C631
Aerobic respiration	3	2.88%	0.02319	Q3S276, B8CAK5, B8C3V9
Cellular macromolecular complex subunit organization	5	4.81%	0.02445	B5YMW7, B8CEQ7, B8BRR6, B8BS24
Monosaccharide catabolic process	4	3.85%	0.03666	B8BXS1, B5YN92, B8C0L7, B8C631
Cellular homeostasis	5	4.81%	0.04037	B8CCL3, B8C4S9, B8LC90, B8LBI2, B8BXF4
Proteolysis involved in cellular protein catabolic process	5	4.81%	0.04734	B8CFM9, B8LBU0, B8LDK1, B8BPY8, B8LCY9
Cellular macromolecule catabolic process	5	4.81%	0.05298	B8CFM9, B8LBU0, B8LDK1, B8BPY8, B8LCY9

Table 1. Biological Enrichment of Gene Ontology (GO) Process Terms in Fe-replete Plastids. Biological enrichment analysis was completed to determine if the subset of proteins determined to be more abundant in Fe-replete plastid fractions (*i.e.*, QSpec results: 105 proteins) represented an enrichment in GO terms relative to GO terms from all proteins identified (929 proteins). Count: number of proteins identified to be from listed GO term; %: percent of protein from this GO term represented in Fe-replete fraction relative to all terms identified in plastid fractions; *p*-value: a modified Fisher Exact *p*-value for gene-enrichment analysis, ranges from 0 to 1. Fisher Exact *p*-value = 0 represents perfect enrichment. *p*-value ≤ 0.05 considered to be strongly enriched in the annotation categories.

Biological Process GO Category	Count	%	*p*-value	UNIPROT Protein IDs
DNA packaging	6	5.0847%	0.03234	B5YMD6
RNA biosynthetic process	3	2.5424%	0.03857	A0T0R0, A0T0Q9, A0T0Z4

Table 2. Biological Enrichment of Gene Ontology (GO) Process Terms in Fe-limited Plastids. Biological enrichment analysis was completed to determine if the subset of proteins determined to be more abundant in Fe-limited plastid fractions (*i.e.*, QSpec results: 128 proteins) represented an enrichment in GO terms relative to GO terms from all proteins identified (929 proteins). Count: number of proteins identified to be from listed GO term; %: percent of protein from this GO term represented in Fe-limited fraction relative to all terms identified in plastid fractions; *p*-value: a modified Fisher Exact *p*-value for gene-enrichment analysis, ranges from 0 to 1. Fisher Exact *p*-value = 0 represents perfect enrichment. *p*-value ≤ 0.05 considered to be strongly enriched in the annotation categories.

limited plastid samples. Using spectral counting statistics from *QSpec* software and thresholds for differential expression (log fold > 0.5 and Z-statistic score > 2) [33, 34]. 128 proteins were found to be significantly more abundant in the Fe-limited samples. The Fe-replete samples had a total of 105 proteins that passed the same strict spectral counting thresholds. Proteins identified by *QSpec* to be either statistically more or less abundant in the different cell states were then examined with respect to all proteins identified to reveal if the subset of more abundant proteins was enriched in any particular biological functions. These enrichment analyses were completed using DAVID software. Examination of the *QSpec* cohorts of differentially expressed proteins did identify enrichments of expressed physiological processes and biochemical pathways. Specifically, DAVID revealed that the Fe-replete plastid protein lists were enriched in gene products involved in the generation of metabolites for energy, monosaccharide processes, catabolism, glycolysis, and aerobic respiration (**Table 1**). Biological enrichment functional summaries of the unique proteins in Fe-limited samples revealed that they were involved in DNA packaging and RNA biosynthetic processes (**Table 2**). Despite more proteins being uniquely identified in the Fe-limited samples (212 vs. 121; **Figure 2**), there were fewer well-represented metabolic pathways in the Fe-limited proteins (**Table 2** vs. **Table 1**). The DAVID software also revealed that both Fe-limited and Fe-replete cultures have biological enrichments of proteins involved in DNA processing, however Fe-limited samples had twice as many unique DNA processing proteins than those of the replete samples (ratio of total proteins associated with DNA processing expressed in Fe+:Fe- plastids: 3:6).

A volcano plot of all proteins identified in these experiments color-coded by location of DNA revealed that chloroplast-originating proteins tend to be present in greater abundance in the Fe-limited cells (**Figure 3**). Analysis of spectral count data from Fe-limited cultures revealed 128 proteins to be more abundant than in Fe-replete culture conditions (**Figure 3**). Although *T. pseudonana* plastid-encoded proteins were observed to have a greater abundance in the Fe-limited cell cultures, the amounts of mitochondrial- and nucleus-encoded proteins were similar in the two growth conditions. No proteins were observed to have above or below 4-fold expressed abundances. The Z-statistic provides a score for the spectral count reproducibility between the sample

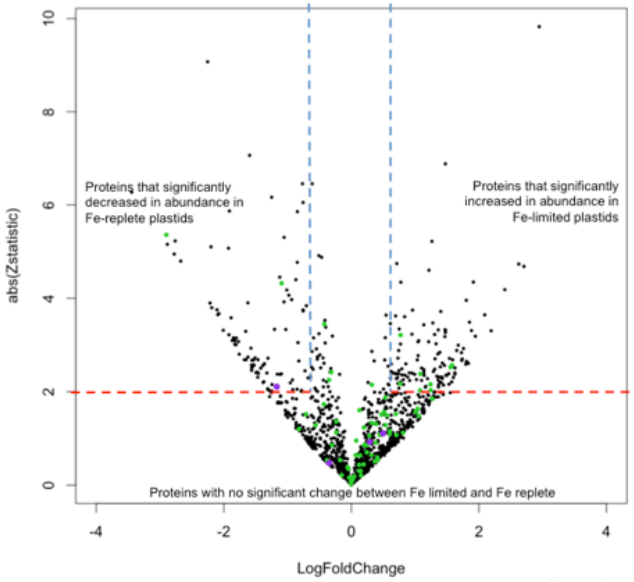

Figure 3. Log fold change vs. Z-statistic scores from *QSpec* analysis summarizing protein expression data of Fe-limited plastid abundances compared to Fe-replete plastid protein abundances. Points are colored according to where their protein sequence is encoded: plastid (green), mitochondrial (purple), or nuclear (black). Positive fold change indicates that a specific protein, represented by a point, had an increased abundance in the Fe-limited plastids relative to the Fe-replete plastids. Blue dashed line indicates the log fold (base 2) change threshold required to be statistically present at higher (positive values > 0.05) or lower abundances (negative values < -0.05). Red dashed lines represent the significance threshold for the absolute Z-statistic score (*i.e.*, > 2 indicate Z-stat score limit that must be passed to be considered statistically different). Using these thresholds for differential expression (log fold > 0.5 and Z-statistic score > 2), 128 proteins were found to be significantly more abundant in the Fe- samples (right upper quadrant), and Fe- replete samples had a total of 105 proteins that were more abundant (left upper quadrant).

suites, where high values indicate samples in which the protein spectral counts between replicates for one cell state did not vary, but as a whole were significantly different from the protein spectral counts for the other cell state. Many of the proteins exhibiting high Z-statistic scores yielded ambiguous annotations (*i.e.*, uncharacterized protein), revealing protein targets to investigate with other molecular techniques. There were 6 proteins with > 2-fold difference in abundance; this represents a suite of proteins that are 2^2 times (*i.e.*, 4 times) more abundant in Fe-limited plastids. These highly abundant proteins in the Fe-limited plastids include 2 cell-surface/membrane proteins and proteins involved in the management and cleavage of DNA. On the other hand, 14 proteins were observed at significantly decreased abundance in the Fe-limited plastids compared to Fe-replete plastids and include proteins involved in amino acid degradation and the Fe-responsive proteins aconitase and dihydrolipoyl dehydrogenase. Additionally, proteins involved in the transfer, catabolism, and conversion of phosphates are within this group of proteins that are present in Fe-limited cells at lower abundances.

Discussion

Here, we set out to determine if there were detectable differences in proteins observed within the plastid of a diatom when grown under Fe-replete and Fe-limited conditions. Our hypothesis that diatoms express different proteins to adapt to Fe-specific environmental stress and that this expression is observable using mass spectrometry was confirmed through these experiments. Using a cluster analysis, a statistical difference was observed in the protein abundance data present in the plastids because of the available Fe (**Figure 1**). Microscopic visual inspection of the collected plastid fraction revealed that the suspended organellar fraction within the Percoll gradient consisted of only chloroplasts. Additionally, comparisons of protein lysates from these whole cell lysates and the plastid-isolated fractions were visualized on a silver-stained SDS-PAGE protein gel [21], revealing an enrichment of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) subunit proteins in the plastid fraction. Furthermore, recent discussions within the molecular technology community have communicated that proteomic mass spectrometry is a valuable method for determining the purity of fractionated cells [15, 16]. For example, examination of all proteins identified in these plastid fractions (929 proteins) with respect to all genes predicted for *Thalassiosira pseudonana* (total of 11,718 proteins in UniProt) revealed significant biological enrichment of the chloroplast and plastid cellular component. Again, there are 127 total proteins that are encoded in the *Thalassiosira pseudonana* chloroplast [12]. Of the 929 proteins identified here, 86 proteins of the 127 proteins were annotated to be from the plastid genome. There were 60 proteins identified in these plastid fractions that are involved specifically in photosynthesis. That said, there are, and a high number of proteins identified within this study to be present within the cleaned, isolated, and microscopically examined plastids from these two diatom cultures that were not translated from the plastid genome. Specifically, there are hundreds of proteins that were nuclear encoded (**Figure 3**), and many of these proteins yielded significantly different abundances, thereby driving the observed statistical differences in the NMDS cluster analysis (*e.g.*, aconitase). Although the presence of these proteins in the plastid fraction initially raised concerns about other-than-organelle contamination, this first look at the diatom plastid proteome reveals high similarity to those plastid proteomes previously published [17] and publically available plastid proteome

databases (http://ppdb.tc.cornell.edu).

From our cluster analysis of all proteomic data, GroEL chaperones appeared to be an important factor driving the differences observed in **Figure 1**. The decreased abundance of GroEL chaperones in Fe-limited cell growth (log fold -2.9; ~8 times more abundant compared to Fe-replete cells) likely negatively affected growth. GroEL chaperones refold unfolded polypeptides, and the lack of GroEL would yield misfolded proteins. A major role for GroEL chaperones in chloroplasts is to help RuBisCO fold properly. It may be that a decrease in photosynthesis under Fe-limitation results in less new RuBisCO being synthesized and a corresponding down-regulation of GroEL. Fe-limited diatom plastids also exhibited a lower abundance of proteins involved in the transfer, catabolism, and conversion of phosphates. The Fe-limited cells were not dividing as rapidly due to the lack of Fe and likely did not have the typical high phosphate demand required for DNA synthesis [18, 19]. Other proteins that were significantly more abundant in the Fe-replete samples and contributed to the separation on the NMDS plot included multiple uncharacterized proteins, representing new areas of potential research now that these can be isolated within the organelle. Additionally, many of the proteins observed within the plastid proteome that are not traditionally considered to be present in the plastid (*e.g.*, mitochondrial glycine decarboxylase P-protein GdcP B8BX31 and aconitase AcnB, B83CV9) were present and also discovered in the chloroplasts of *Arabidopsis thaliana* [20]. This suggests that the high number of unique, non-plastid-encoded proteins found is in agreement with previous research on terrestrial plants. That said, some of these non-plastid-encoded proteins are possibly due to contamination from the rest of the cell, even though it appeared that plastid fractions were clean [21].

As diatoms are key contributors to the global carbon and silica cycle [22, 23], the effects of Fe on general photosynthetic capacity are important to understand. Here, we revealed that Fe directly controls the proteins expressed within the plastid, the organelle that houses the reactions for photosynthesis. Diatoms require Fe as a cofactor to transfer electrons toward the ATP synthase complex; Fe is essential to the mechanics of Photosystem I (PSI), PSII, and cytochrome processes [24]. The data resulting from our analysis of *T. pseudonana* grown under Fe-limited conditions demonstrates that the cells must alter the plastid proteome to adapt to diminished Fe concentrations. Further, we show that *T. pseudonana* is able to adapt to low Fe conditions and maintain the core complexes of PSI and PSII, but the majority of the subcomplexes are down regulated. Therefore, the photosynthetic capacity of diatoms was challenged when Fe was limited, but it was not catastrophic since growth was observed, albeit at low rates. This is further evidence that Fe-limitation leads to a unique phenotypic expression for survival. Through the isolation and analysis of plastid proteins, as shown here, scientists can now better identify what nutrients and/or trace metals directly affect diatom photosynthetic capacity so they can design better experiments to increase CO_2 fixation rates.

Materials and Methods

Diatom Cultures.

The diatom *Thalassiosira pseudonana* was selected as the study organism as it has a well-characterized physiology, genome, and proteome [11, 25-27]. *T. pseudonana* clone 3H (CCMP1335, mean diameter: 4 µm) was obtained from the Provasoli-Guillard Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME, USA). Cells were grown axenically using semi-continuous batch culturing with a modified version of f/2 made in 0.2-µm filtered and microwave-sterilized Sargasso seawater [28]. Fe was added separately to achieve the desired concentrations. Biological duplicate cultures grown under replete conditions received 400 nM Fe. The inoculum for the biological duplicate Fe-limited treatment came from an exponentially growing replete culture that was diluted (1:10) in media without added Fe. Cultures were grown in 2-L acid-cleaned polycarbonate flasks. All cultures were maintained at 13°C, at a continuous light intensity of 150 µmol quanta m^{-2} s^{-1}, and provided with cool-white fluorescent lights. The growth rates (d^{-1}) of the cultures were monitored daily using *in vivo* chlorophyll a (Chl) fluorescence measurements with a Turner Designs Trilogy Fluorometer (Sunnyvale, CA, USA). Cell density (cells per mL) and cell counts were completed using a haemocytometer with an Olympus BX-50 light microscope. Sterile trace metal–clean techniques were used during all experiments and manipulations.

Chloroplast and Other Plastid Extraction from Diatoms.

Diatom cultures were grown in autoclaved f/2 media until the cell density reached approximately 1E6 cells per ml. Two samples were isolated from biological duplicate cultures from each treatment (Fe replete and Fe limited). Due to the limited number of cells in the Fe-limited cultures, only three distinct samples were isolated from the biological duplicate cultures for proteomics (two culture flasks, where one flask yielded two samples for mass spectrometry [MS] analyses). Two splits were generated from each biological duplicate flask of the Fe-replete samples, yielding a total of four samples for MS analyses. The protocol to isolate organelles from diatoms was followed to yield a clean plastid-only fraction [21]. Isolation buffer consisted of a mixture of sorbitol, Na_2-EDTA, $MgCl_2$, KCl, $MnCl_2$, HEPES-KOH (pH = 8.0), and

sterile H_2O. Just before the isolation buffer was used, 1% (w/v) bovine serum albumin (BSA) was added. PEG-6000 solution was made by combining poly-ethylene-glycol (PEG) with sterile H_2O. The Percoll solution was generated from Percoll, PEG-6000 solution, Ficoll, and BSA. The gradient mixture combined HEPES-KOH pH = 8.0, EDTA, sorbitol, and sterile H_2O. The Percoll gradient had a bottom layer of 80% Percoll and a top layer of 40% Percoll. A small amount of gel loading dye was added to the 40% Percoll solution before it was gently placed above the bottom layer Percoll solution using a Pasteur pipette

Pelleted diatom cells had their silica cell walls compromised using NH_4F and vortexing. The slurry was then centrifuged, and overlying NH_4F was poured off. Cells were resuspended and rinsed 3x with isolation buffer and centrifuged between rinses. Cells were then resuspended in a small amount of isolation buffer and briefly sonicated at the lowest setting for 10s. The diatom homogenate was loaded carefully onto the top of each two-step Percoll gradient with a Pasteur pipette. Cells were then centrifuged in a swing-out rotor with the brake off. Intact chloroplasts were visible between the layers and were recovered using a Pasteur pipette. The chloroplasts were then washed twice with isolation buffer without added BSA. To wash, buffer was added and chloroplasts were centrifuged in a swing-out rotor (1000xg, 5 min, 4°C), and the supernatant was discarded. Chloroplast fraction was then interrogated using phase contrast microscopy to visualize and confirm successful extraction of intact chloroplasts. Additionally, chloroplast fractions were run on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) systems to determine purity [21].

Protein Sample Preparation and Mass Spectrometry. Sample preparation and mass spectrometry of plastid fractions for each treatment were completed following Nunn *et al*. [10] using an Orbitrap mass spectrometer. Plastid samples were ruptured using a sonicating probe, and lysates were then solubilized with 100 µl 6M urea. The pH was raised to 8.8 by adding 6.6 µl of 1.5 M Tris HCl prior to adding 2.5 µl of 200 mM TCEP. This mixture was incubated at 37°C for 1 hr. Proteins were then alkylated with the addition of 20 µl of 200 mM iodoacetamide (IAM), vortexed, and incubated for 1 hr at room temperature in the dark. Next, 20 µl of 200 mM dithiolthreitol (DTT) was added to the mixture, vortexed, and incubated 1 hr to soak up excess IAM. Prior to the addition of trypsin, the urea mixture was diluted by adding 800 µl of 25 mM NH_4HCO_3 and 200 µl MeOH (HPLC grade). Trypsin was added to the mixture at a ratio of 30 protein:1 trypsin and incubated overnight at 37°C. Samples were then speedvaced to near dryness (~20ul) and desalted to remove urea salts following manufacturers' protocols (C18 macrospin column; Nest Group Inc.) Desalted samples were speedvaced to ~1 µl and resuspended in 0.1% formic acid and 5% ACN to a final concentration of 1 µg/µl. Samples were stored at -80°C prior to analysis.

Resulting peptide samples were separated using inline reverse-phase chromatography (20 cm long, 75 µm i.d. fused silica capillary column packed with C18 particles fitted with a 2 cm long, 100 µm i.d. precolumn). Peptides were eluted using an acidified water-acetonitrile gradient (5–35% ACN, 0.1% formic acid). Mass spectrometry was performed on a Thermo Fisher hybrid tandem mass spectrometer, the linear ion trap Orbitrap (LTQ-OT). A total of 1 µg of peptide digest was sampled per LC-MSMS analysis.

All tandem mass spectrometry results were searched and interpreted with Comet and then scored with the Trans-Proteomic-Pipeline (TPP) [29-31]. The protein database used for correlating spectra with peptides consisted of the latest release version 3.0 of the *T. pseudonana* predicted protein database (www.jgi.doe.gov), the unmapped sequences (Thaps3_bd; www.jgi.gov), bovine serum albumin sequence, and 50 common contaminants. Comet parameters included reverse concatenated sequence database search, trypsin enzyme specificity, and variable modifications on cysteine (57 Da) and methionine (15.999 Da: oxidation). Minimum protein and peptide thresholds were set at $p > 0.95$ on ProteinProphet and PeptideProphet [29]. Using concatenated target-decoy database searches, false-discovery rates (FDR) were calculated according to Elias and Gygi [32] and were all <1%.

Label-free Protein Quantification (QSpec). For protein quantification, two splits from biological culture replicates were collected from the Fe-limited condition (total of four samples). We also collected two splits from one biological replicate plus one split from a second biological replicate from Fe-replete cultures (total of three samples). Each of the 7 samples received two 90-minute analyses on the LTQ-OT (total of 14 analyses on the MS), representing technical replicates. This dataset was used to determine the relative quantities of protein expression from the Fe-replete and Fe-limited cultures. Spectral counting was employed to determine relative quantities [33, 34]. Spectral counting data were filtered at $p > 0.95$ protein and peptide probability using PeptideProphet, and proteins were only considered if > 2 peptides were identified in each technical and biological replicate. Significance analyses were completed using *QSpec* to determine if proteins were significantly more or less abundant. *QSpec* is reported using a fold change difference in abundance with a log base 2 scale;

therefore, a log fold change of 1 indicates it is twice as abundant (2^1). Proteins were only considered to be present at significantly different abundance levels if the reported Z-score was ≥ 2 or ≤ -2 and the corresponding fold difference observed was ≥ 0.5 or ≤ -0.5.

R Statistics. To assess if the proteins identified in Fe-replete bioreplicates and Fe-limited bioreplicates were statistically different, QSpec data was analyzed on a non-metric multidimensional scaling (NMDS) plot [35, 36]. NMDS plots of all MS technical replicate proteomic data were completed to determine if there were significant differences within analyses of biological replicate cultures or biological splits from the same culture (not shown). Once this was completed, spectral counts for technical replicates were averaged and then biological replicates or splits were plotted (**Figure 1**). Volcano plots were created using all QSpec data and cellular location data based on Joint Genome Institute predictions as entered into the UniProt genome server.

Biological Enrichment Analyses. Significant differences in protein abundance were determined using *QSpec*. These proteins represent the important physiological processes that are responding to Fe-limitation by either increasing or decreasing abundances. To determine if these sets of differential proteins were specific to a biological pathway or function, an unbiased enrichment analysis was completed. The Database for Annotation, Visualization and Integrated Discovery (DAVID; v6.7) was used to identify biological enrichments within the context of the larger dataset that included all proteins identified [11, 37]. Proteins that were determined to be at significantly higher or lower abundance in response to Fe-limitation were examined by this functional annotation tool in the context of a background expressed proteome (all proteins identified across biological and technical replicates). Specifically, using DAVID's functional annotation tool, the background gene list included all proteins identified in the plastid fractions, and the gene list under investigation was either the list of significantly increased or decreased protein abundance in the Fe-limited cells (as determined by *QSpec*).

Acknowledgements

This work was supported and funded by a grant from the National Science Foundation (NSF-OCE 1233014) for BLN, SJN, and AB, National Science Foundation Grants OCE 0962208 and PLR 1443474) to BDJ as well as a Training Grant from the National Institutes of Health for the lab of BLN (T32 HG00035). We acknowledge support from National Science Foundation Grant OCE-1060300 and support in part by the University of Washington Proteomics Resource Bioinformatics Team Jimmy Eng and Priska von Haller (UWPR95794).

References

1. Greene, R. M., R. J. Geider, and P. G. Falkowski. "Effect of Iron Limitation on Photosynthesis in a Marine Diatom." *Limnology and Oceanography* 36.8 (1991): 1772-82.
2. Behrenfeld, M. J., *et al*. "Confirmation of Iron Limitation of Phytoplankton Photosynthesis in the Equatorial Pacific Ocean." *Nature* 383.6600 (1996): 508-11.
3. Boyd, P. W., *et al*. "A Mesoscale Phytoplankton Bloom in the Polar Southern Ocean Stimulated by Iron Fertilization." *Nature* 407.6805 (2000): 695-702.
4. Buesseler, K. O., *et al*. "The Effects of Iron Fertilization on Carbon Sequestration in the Southern Ocean." *Science* 304.5669 (2004): 414-7.
5. Sarmiento, J. L., *et al*. "Efficiency of Small Scale Carbon Mitigation by Patch Iron Fertilization." *Biogeosciences* 7.11 (2010): 3593-624.
6. Hopwood, M. J., *et al*. "Glacial Meltwater from Greenland Is Not Likely to Be an Important Source of Fe to the North Atlantic." *Biogeochemistry* 124.1-3 (2015): 1-11.
7. Buesseler, K. O., *et al*. "Environment. Ocean Iron Fertilization--Moving Forward in a Sea of Uncertainty." *Science* 319.5860 (2008): 162.
8. Nelson, D. M., *et al*. "Production and Dissolution of Biogenic Silica in the Ocean - Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation." *Global Biogeochemical Cycles* 9.3 (1995): 359-72.
9. Lommer, M., *et al*. "Genome and Low-Iron Response of an Oceanic Diatom Adapted to Chronic Iron Limitation." *Genome Biology* 13.7 (2012): R66.
10. Nunn, B. L., *et al*. "Diatom Proteomics Reveals Unique Acclimation Strategies to Mitigate Fe Limitation." *PLoS ONE* 8.10 (2013): e75653.
11. Armbrust, E. V., *et al*. "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism." *Science* 306.5693 (2004): 79-86.
12. Oudot-Le Secq, M. P., *et al*. "Chloroplast Genomes of the Diatoms Phaeodactylum Tricornutum and Thalassiosira Pseudonana: Comparison with Other Plastid Genomes of the Red Lineage." *Mol Genet Genomics* 277.4 (2007): 427-39.
13. Grouneva, I., A. Rokka, and E. M. Aro. "The Thylakoid Membrane Proteome of Two Marine Diatoms Outlines Both Diatom-Specific and Species-Specific Features of the Photosynthetic Machinery." *Journal of Proteome Research* 10.12 (2011): 5338-53.
14. Gruber, A., *et al*. "Plastid Proteome Prediction for Diatoms and Other Algae with Secondary Plastids of the Red Lineage." *Plant Journal* 81.3 (2015): 519-28.
15. Aebersold, R., and M. Mann. "Mass Spectrometry-

Based Proteomics." *Nature* 422.6928 (2003): 198-207.
16. Aebersold, R., A. L. Burlingame, and R. A. Bradshaw. "Western Blots Versus Selected Reaction Monitoring Assays: Time to Turn the Tables?" *Molecular & Cellular Proteomics* 12.9 (2013): 2381-82.
17. Bischoff, R., and H. Schluter. "Amino Acids: Chemistry, Functionality and Selected Non-Enzymatic Post-Translational Modifications." *J Proteomics* 75.8 (2012): 2275-96.
18. Marchetti, A., and P. J. Harrison. "Coupled Changes in the Cell Morphology and the Elemental (C, N, and Si) Composition of the Pennate Diatom Pseudo-Nitzschia Due to Iron Deficiency." *Limnology and Oceanography* 52.5 (2007): 2270-84.
19. Boyd, P. W., *et al*. "Biological Ramifications of Climate-Change-Mediated Oceanic Multi-Stressors." *Nature* Climate Change 5.1 (2015): 71-79.
20. Kleffmann, T., *et al*. "The Arabidopsis Thaliana Chloroplast Proteome Reveals Pathway Abundance and Novel Protein Functions." *Current Biology* 14.5 (2004): 354-62.
21. Chappell, P. Dreux, and Bethany D. Jenkins. "Chloroplast Isolation from Diatoms." www.protocols.io 2016.
22. DiTullio, G. R., *et al*. "Rapid and Early Export of Phaeocystis Antarctica Blooms in the Ross Sea, Antarctica." *Nature* 404.6778 (2000): 595-98.
23. Falkowski, P. "Ocean Science: The Power of Plankton." *Nature* 483.7387 (2012): S17-20.
24. Raven, J. A., M. C. W. Evans, and R. E. Korb. "The Role of Trace Metals in Photosynthetic Electron Transport in O-2-Evolving Organisms." *Photosynthesis Research* 60.2-3 (1999): 111-49.
25. Nunn, B. L., *et al*. "Deciphering Diatom Biochemical Pathways Via Whole-Cell Proteomics." *Aquatic Microbial Ecology* 55.3 (2009): 241-53.
26. Carvalho, R. N., and T. Lettieri. "Proteomic Analysis of the Marine Diatom Thalassiosira Pseudonana Upon Exposure to Benzo(a)Pyrene." *BMC Genomics* 12 (2011): 159.
27. Dyhrman, S. T., *et al*. "The Transcriptome and Proteome of the Diatom Thalassiosira Pseudonana Reveal a Diverse Phosphorus Stress Response." *PLoS ONE* 7.3 (2012): e33768.
28. Guillard, R. R. L., and P. E. Hargreaves. "Stichochrysis-Immobilis Is a Diatom, Not a Chrysophyte (Vol 32, Pg 234, 1993)." *Phycologia* 33.1 (1994): 66-66.
29. Keller, A., *et al*. "Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by Ms/Ms and Database Search." *Analytical Chemistry* 74.20 (2002): 5383-92.
30. Eng, J. K., T. A. Jahan, and M. R. Hoopmann. "Comet: An Open-Source Ms/Ms Sequence Database Search Tool." *Proteomics* 13.1 (2013): 22-4.
31. Nesvizhskii, A. I., *et al*. "A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry." *Anal Chem* 75.17 (2003): 4646-58.
32. Elias, J. E., and S. P. Gygi. "Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry." *Nature Methods* 4.3 (2007): 207-14.
33. Choi, H., D. Fermin, and A. I. Nesvizhskii. "Significance Analysis of Spectral Count Data in Label-Free Shotgun Proteomics." *Mol Cell Proteomics* 7.12 (2008): 2373-85.
34. Old, W. M., *et al*. "Comparison of Label-Free Methods for Quantifying Human Proteins by Shotgun Proteomics." *Mol Cell Proteomics* 4.10 (2005): 1487-502.
35. Oksanen, Jari, *et al*. Vegan: Community Ecology Package. R Package. Vers. 2013.
36. Team, R Core. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vers. 2013.
37. Huang DW, Sherman BT, Lempicki RA. "Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources." *Nature Protoc* 4.1 (2009): 44-57.