10-1-2023

Subcellular proteomics for determining iron-limited remodeling of plastids in the model diatom Thalassiosira pseudonana (Bacillariophyta)

Kristofer M. Gomes
University of Rhode Island

Brook L. Nunn
University of Washington

P. Dreux Chappell
University of South Florida St. Petersburg

Bethany D. Jenkins
University of Rhode Island

Citation/Publisher Attribution

Gomes, Kristofer M., Brook L. Nunn, P. D. Chappell, and Bethany D. Jenkins. "Subcellular proteomics for determining iron-limited remodeling of plastids in the model diatom Thalassiosira pseudonana (Bacillariophyta)." *Journal of Phycology* 59, 5 (2023). doi: 10.1111/jpy.13379.

Subcellular proteomics for determining iron-limited remodeling of plastids in the model diatom Thalassiosira pseudonana (Bacillariophyta)

Keywords

chloroplast; diatom; iron; photosynthesis; plastid; proteome

Received: 13 March 2023 | Accepted: 14 July 2023

DOI: 10.1111/jpy.13379

RESEARCH ARTICLE

Subcellular proteomics for determining iron-limited remodeling of plastids in the model diatom *Thalassiosira pseudonana* (Bacillariophyta)

Kristofer M. Gomes[1] | **Brook L. Nunn**[2] | **P. Dreux Chappell**[3] | **Bethany D. Jenkins**[4,5]

[1]Department of Biological Sciences, University of Rhode Island, Rhode Island, Kingston, USA

[2]Department of Genome Sciences, University of Washington, Washington, Seattle, USA

[3]College of Marine Science, University of South Florida, Florida, St. Petersburg, USA

[4]Department of Cell and Molecular Biology, University of Rhode Island, Rhode Island, Kingston, USA

[5]Graduate School of Oceanography, University of Rhode Island, Rhode Island, Narragansett, USA

Correspondence
Kristofer M. Gomes, Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI 02881, USA.
Email: kristofer_gomes@uri.edu

Funding information
National Science Foundation, Grant/Award Number: OCE 1633939, OCE 1756816 and OPP 144374; University of Washington, Grant/Award Number: UWPR95794

Editor: N. Poulsen

Abstract

Diatoms are important primary producers in the world's oceans, yet their growth is constrained in large regions by low bioavailable iron (Fe). Low-Fe stress-induced limitation of primary production is due to requirements for Fe in components of essential metabolic pathways including photosynthesis and other chloroplast plastid functions. Studies have shown that under low-Fe stress, diatoms alter plastid-specific processes, including components of electron transport. These physiological changes suggest changes of protein content and in protein abundances within the diatom plastid. While in silico predictions provide putative information on plastid-localized proteins, knowledge of diatom plastid proteins remains limited in comparison to well-studied model photosynthetic organisms. To address this, we employed shotgun proteomics to investigate the proteome of subcellular plastid-enriched fractions from *Thalassiosira pseudonana* to gain a better understanding of how the plastid proteome is remodeled in response to Fe limitation. Using mass spectrometry-based peptide identification and quantification, we analyzed *T. pseudonana* grown under Fe-replete and -limiting conditions. Through these analyses, we inferred the relative quantities of each protein, revealing that Fe limitation regulates major metabolic pathways in the plastid, including the Calvin cycle. Additionally, we observed changes in the expression of light-harvesting proteins. In silico localization predictions of proteins identified in this plastid-enriched proteome allowed for an in-depth comparison of theoretical versus observed plastid-localization, providing evidence for the potential of additional protein import pathways into the diatom plastid.

KEYWORDS
chloroplast, diatom, iron, photosynthesis, plastid, proteome

Abbreviations: ER, endoplasmic reticulum; FDR, false discovery rate; Fe, Iron; GO, Gene Ontology.

INTRODUCTION

Diatoms are important primary producers in the world's oceans, generating approximately 40% of the organic carbon fixed annually (Nelson et al., 1995). Despite their role as primary producers within the marine environment, their rates of growth and primary production are constrained in large regions of the surface ocean by low iron (Fe) availability (Geider & La Roche, 1994; Moore et al., 2013; Sunda & Huntsman, 1995). The limitation of diatom growth and primary production from Fe insufficiency is a result of the biochemical requirement for Fe, an important co-factor in essential metabolic pathways, including key plastid-localized pathways such as photosynthesis and nitrogen metabolism (Geider & La Roche, 1994; Milligan & Harrison, 2000). Following the introduction of Fe to Fe-limited waters, diatom populations bloom (Martin et al., 1990; Sunda & Huntsman, 1995), indicating phenotypic plasticity that allows for both their survival in low Fe environments and rapid growth when Fe becomes more abundant. Prior studies have shown that under Fe-limiting conditions, diatoms alter plastid-specific characteristics, including protein components of the electron transport chain, chlorophyll content, and light-harvesting complexes, suggesting plastid-specific protein alterations via their photosynthetic activity (Allen et al., 2008; Cohen et al., 2018; Moreno et al., 2017, 2020; Nunn et al., 2013). As diatoms are the foundation of many marine food webs, an improved understanding of the diatom plastid response to low-Fe conditions will provide mechanistic insights into the consequences of Fe insufficiency on marine ecosystems and further insight into the differing Fe metabolism strategies present across diatom species (Gao et al., 2021).

Although plastids, including those from diatoms, maintain their own reduced genomes and protein synthesis machinery, their function ultimately depends on the import of proteins that derive from nuclear-encoded genes (Allen, 2015; Bhattacharya et al., 2007). These nucleus-encoded genes for plastid function were transferred over evolutionary time to the diatom nuclear genome during a series of endosymbiotic events. Primary endosymbiosis occurred from the engulfment of a free-living cyanobacterial endosymbiont by a eukaryotic cell, resulting in a photosynthetic cell that gave rise to all eukaryotic photosynthetic lineages (Archibald, 2015). Following this primary endosymbiotic event, a secondary endosymbiosis occurred during which a primary photosynthetic cell of the red algal lineage was engulfed by a second non-photosynthetic eukaryote (Stiller et al., 2014), giving rise to the evolutionary lineage from which diatoms are derived. Diatom plastid envelope structure, composed of three or four envelope membranes (Kroth, 2002), reflects these serial endosymbiotic events. The outermost membrane is contiguous with the endoplasmic reticulum (ER), with the two most inner membranes homologous to the inner and outer membranes of primary plastids found in plants (Burki et al., 2016).

To reach their plastid destination, proteins encoded by the diatom nucleus require N-terminal sequences that direct transit across the ER and plastid envelope membranes (Kroth, 2002), which are then cleaved by a peptidase in the plastid stroma (Huesgen et al., 2013). Consensus sequence analysis of diatom nucleus-encoded proteins with plastid function has identified a N-terminal bipartite peptide sequence involved in the transport of proteins across plastid membranes. This N-terminal sequence contains a highly conserved amino acid motif comprised of "ASAFAP" residues upstream of a required phenylalanine residue at the +1 of the cleavage site (Figure 1; Apt et al., 2002; Gruber et al., 2007). The characterization of the amino acid composition of the diatom N-terminal plastid targeting motif has facilitated genome-wide computational predictions of nuclear-encoded proteins that are likely to be imported into diatom plastids (Gruber et al., 2015; Gschloessl et al., 2008).

Biochemical confirmation of proteins functioning in diatom plastids is the next step for understanding the physiology of these organelles essential for diatom function and primary production. In the plant model

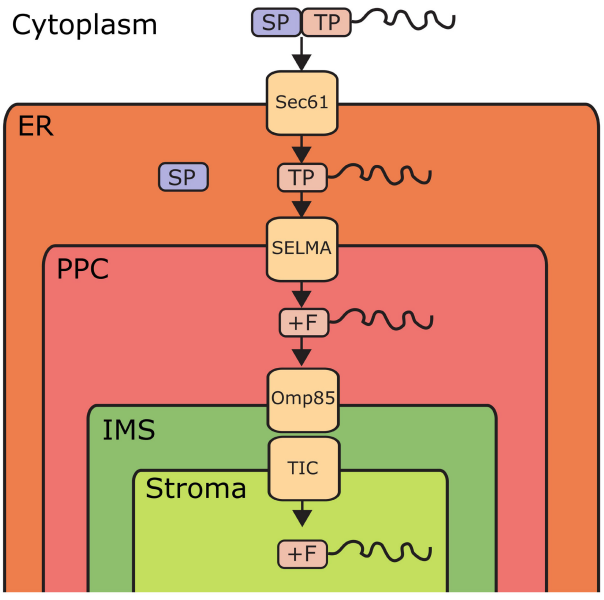

FIGURE 1 Canonical plastid import pathway of nuclear-encoded proteins. Import is dependent on the presence of bipartite transit peptide, composed of a signal (SP) and transit peptide (TP; Gruber et al., 2007; Hempel et al., 2009). Transport across the endoplasmic reticulum (ER) membrane is performed via Sec61, dependent on the SP. Transport into the periplasmic compartment (PPC), dependent on the TP, is performed via symbiont-specific ERAD-like machinery (SELMA; Stork et al., 2012). Further transport into the plastid stroma is dependent on the presence of phenylalanine (F) at the +1 position of the TP and is performed through the Omp85/TIC complex. [Color figure can be viewed at wileyonlinelibrary.com]

species *Arabidopsis thaliana*, plastid isolation, in combination with proteome sequencing, has identified a number of proteins not computationally predicted to be targeted to the primary plastid (Friso et al., 2004; Kleffmann et al., 2004; Wang et al., 2016). This gap between computational prediction and plastid protein composition in a well-characterized plant model system suggests that organellar protein import and trafficking are more complex than was previously anticipated (Kleffmann et al., 2004). Profiling the proteins in plastids of diatoms, eukaryotes with a more complex plastid membrane architecture than plants, will be important for elucidating diatom plastid function and structure.

To investigate the protein content localized to the plastids of diatoms, we purified intact diatom plastids in the model centric diatom species *Thalassiosira pseudonana* and analyzed the lysates with mass spectrometry (MS)-based shotgun proteomics. In addition to aiding in the elucidation of the diatom plastid-localized proteome, this study compared plastid-enriched proteomes collected from *T. pseudonana* cultures grown in both Fe-replete and Fe-limiting conditions to identify how the plastid proteome remodels under Fe-limitation. While previous analyses have shown that *T. pseudonana* has a limited inventory of canonical Fe-sensitive genes, including the lack of an Fe-sensitive isoform of flavodoxin (Graff van Creveld et al., 2023; Morrissey et al., 2015; Whitney et al., 2011), it was chosen for this study as it has a well-characterized Fe-limited transcriptome and whole-cell proteome (Mock et al., 2008; Nunn et al., 2013), a fully sequenced genome (Armbrust et al., 2004), and a plastid-targeted proteome containing 217 proteins (Schober et al., 2019). Within this study, we report the identification of 929 proteins within the plastid-enriched proteome of *T. pseudonana* across both Fe-limiting and Fe-replete conditions, expanding the inventory of proteins localized within the *T. pseudonana* plastid proteome. Between these datasets, we have observed statistically significant changes in the abundance of 135 proteins in response to changes in Fe levels. In addition, by combining our dataset with computational localization prediction, we have determined an inventory of proteins predicted within distinct regions of the plastid, and those proteins associated with the organelle under replete and Fe-limiting conditions.

An understanding of the metabolic responses of diatoms to differing Fe concentrations in the marine environment has become increasingly important, with predicted increases in terrestrial and anthropogenic sources of Fe that may alter nutrient co-limitation dynamics and alter plankton community structure (Hutchins & Boyd, 2016). Thus, an understanding of diatom responses to differing Fe regimes is a crucial factor in understanding future impacts of climate change on the community structure, nutrient utilization, and output of these important primary producers.

MATERIALS AND METHODS

Cell culturing

Thalassiosira pseudonana (CCMP 1335) were grown under Fe-replete and Fe-limiting f/2 media (Guillard & Ryther, 1962) utilizing trace metal clean procedures. Cultures were maintained in f/2 media made using filter-sterilized Sargasso seawater amended with trace metal clean nutrient stocks with Fe-replete media receiving an additional 400 nM Fe. Fe-limiting f/2 media contained no added Fe. As the Fe-limited culture was initiated with two successive 10-fold dilutions of cultures grown in Fe-replete media, we estimate it to have a maximum of 4 nM Fe. All cultures were maintained at 13°C, at a continuous light intensity of 150 μmol quanta $\cdot m^{-2} \cdot s^{-1}$. The growth of cultures was assessed using duplicate cell counts. Individual cultures were successively cultured under each respective Fe condition (Fe-limiting: 34 d, Fe-replete: 28 d) until growth rates were normalized across successive subcultures to ensure acclimation of *T. pseudonana* cells to each condition (Table S1 in the Supporting Information). Prior to plastid enrichment, single cultures of *T. pseudonana* from Fe-condition acclimated cultures were performed for 4 d under both Fe-replete and Fe-limiting conditions (Table S1). Growth rates of cultures were assessed, with Fe-replete cultures exhibiting over double the specific growth rate relative to the Fe-limiting condition (Fe-replete: 0.54, Fe-limiting: 0.22). On the day of plastid enrichment and protein purification, the Fe-replete culture reached a cell density of ~1.54E+06 cells $\cdot mL^{-1}$ compared to ~1.1054E+06 cells $\cdot mL^{-1}$ under Fe-limiting conditions.

Plastid organelle enrichment

Following culture acclimation to their respective Fe conditions, each culture was divided into triplicate cultures during the exponential growth phase. Subsequently, plastid enrichment was carried out, resulting in three technical replicates for each Fe condition. To prepare a plastid-enriched fraction from cultures, cells were harvested by centrifugation and resuspended in a minimal amount of f/2 media. To collect intact organelles, the diatom silica frustule was dissolved using ammonia fluoride (10 M; Chappell & Jenkins, 2017). Cells were then sonicated briefly (10 s, 10 W) to disrupt the plasma membrane while maintaining organelle integrity (Chappell & Jenkins, 2017). Intact plastid-enriched fractions were then collected using a 40/80 percoll density gradient (Chappell & Jenkins, 2017). Enrichment of the plastid fraction was assessed through SDS-PAGE. Plastid enrichment was assessed through visual inspection of protein bands at ~58 kDA, corresponding to the large subunit of RubisCO (RbCL; ~51–58 kDa; Figure 2).

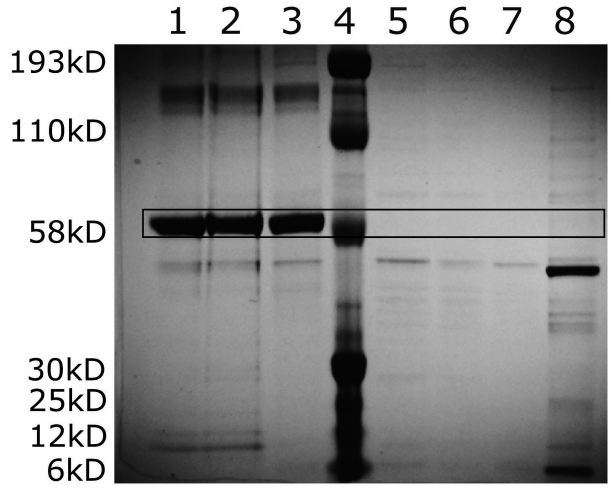

FIGURE 2 SDS-PAGE gel of diatom whole cell and plastid-enriched fractions. The order of samples on the example protein gel are as follows: 1–*Thalassiosira pseudonana* plastid +Fe, 2–*Thalassiosira pseudonana* plastid –Fe, 3–*Thalassiosira weissflogii plastid* +Fe, 4–protein ladder, 5–*Thalassiosira pseudonana* whole cell +Fe, 6–*Thalassiosira pseudonana* whole cell –Fe, 7–*Thalassiosira weissflogii* whole cell +Fe, 8–Spinach whole cell.

Note that the gels were only completed for visualization of the proteins, not for downstream mass spectrometry-based proteomics analyses. Enrichment of the plastid fraction was observed through the visual inspection of bands. Plastids and whole cells were extracted by ice-cold homogenization (Barkan, 1993). Five micrograms of each sample was mixed with Invitrogen™ LDS buffer and β-mercaptoethanol, denatured at 70°C, and run on a 4%–12% Bis-Tris mini gel with MOPS buffer, with Bio-Rad's Kaleidoscope Prestained Protein Standard. Gels were stained with Invitrogen™ SimplyBlue™ SafeStain and imaged on a Syngene Genius gel documentation system.

Protein sample prep and mass spectrometry

Plastid-enriched samples were sonicated in 6 M urea in 50 mM ammonium bicarbonate to solubilize proteins prior to in-solution reduction, alkylation, and trypsin digestion (details found in Nunn et al., 2013). The Pierce BCA Protein Kit microplate assay (Thermo Scientific) was used, following the manufacturer's instructions, to measure protein concentrations. Aliquots of 50 μg of protein sample were subjected to protein digestion. Protein lysates were solubilized in 100 μL of 6 M urea buffer and reduced with 1 μL of 500 mM tris(2-carboxyethyl)phosphine. The pH was then controlled with 6.6 μL of 1.5 M Tris pH 8.8 and adjusted as needed to achieve a pH greater than 8. Next, proteins were alkylated with iodoacetamide (200 mM IAA; 20 μL), and any remaining alkylating agent was quenched with diothiothreitol (200 mM DTT; 20 μL). The urea was diluted with 800 μL of 25 mM NH_4HCO3 and 200 μL HPLC grade methanol before trypsin digestions were carried out. Modified porcine sequencing grade trypsin (Promega) was added in a 1:20 enzyme:sample protein ratio and allowed to cleave proteins overnight at room temperature. The sample pH was then lowered to less than 2 with 10% trifluoroacetic acid (TFA), and peptides were dried at 4°C in a speedvac. Dried plastid peptide samples were reconstituted in 5% acetonitrile and 0.1% trifluoroacetic acid, then desalted using C18 macrospin columns (The NestGroup) according to the manufacturer's instructions. The eluted peptides were dried in a speedvac to a final volume of less than 10 μL and reconstituted in 100 μL of 2% acetonitrile and 0.1% formic acid. The resulting peptides were analyzed on linear ion trap Orbitrap (LTQ-OT) inline with a nanoAquity UPLC chromatography system with an electrospray ionization interface. Peptides were separated using inline reverse-phase chromatography (20 cm, 75 μm i.d. fused silica column and 2 cm, 100 μm i.d. pre-column packed with 3 μm C18 beads Dr. Maisch). For each sample, 1 μg of peptide digest was eluted using a 90-min acidified water acetonitrile gradient (5%–35% ACN, 0.1% formic acid) and analyzed on the LTQ-OT in data-dependent mode, top 20 ions.

Peptides and proteins from *Thalassiosira pseudonana* were identified by searching the resulting raw mass spectrometry files against a concatenated forward and reverse *T. pseudonana* protein sequence database using Comet v.22016.01 (Eng et al., 2013; Friso et al., 2004). The final 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), and 50 common contaminants (cRAPome). Comet parameters included concatenated decoy search, 20 ppm peptide mass tolerance, enzyme set at trypsin with two missed cleavages allowed, variable cysteine modification of 57 Da, and fixed methionine modification of 15.999 Da. After Comet searches were completed, the Trans Proteomic Pipeline was used, and minimum protein and peptide thresholds were set at $p > 0.95$ on ProteinProphet and PeptideProphet, yielding a false discovery rate <1%. Protein identifications from the whole-cell lysates were reported if the above-mentioned thresholds were passed, two or more peptides were identified, and at least one terminus was tryptic (Keller & Shteynberg, 2011). Using concatenated target-decoy database searches, false discovery rates (FDR) were <1%, following FDR calculations by Elias and Gygi (2010).

Protein abundances were assessed using spectral count data with normalization between mass spec analyses, performed using Abacus (Fermin et al., 2011), with differences in protein abundance statistically performed using QSPEC (Choi et al., 2008).

Proteins were considered high confidence if two unique peptides were identified across all mass spectrometry experiments. The threshold for statistically significant changes in abundance of proteins between conditions was defined as a reported BayesFactor > 10, the corresponding FDR <1%, and an absolute value of $\log_2$ fold change > 0.5. Data are available via ProteomeXchange with identifier PXD040824.

KEGG mapping and GO term enrichment analysis

KEGG mapping was performed using BLASTKoala (Kanehisa et al., 2016) for all proteins of the *Thalassiosira pseudonana* plastid-enriched proteome and the UniProt *Arabidopsis thaliana* reference plastid proteome (1163 proteins). Further functional annotation was performed using Eggnog (eggNOG v5.0, eggnog-mapper v2; Huerta-Cepas et al., 2017, 2018). Enrichment of metabolic function within the plastid proteome was performed using PANTHER's (Mi et al., 2013) statistical overrepresentation test, run using default parameters with Bonferroni correction. In addition to analyzing the whole plastid dataset, GO term enrichment was performed on the *T. pseudonana* proteins within the network bins created during sequence homology network analysis.

Computational prediction of plastid proteome localization

Localization prediction for all nuclear-encoded proteins within the dataset was performed using trimmed N-terminal sequences (75 amino acids long). These N-terminal peptide sequences were run subsequently through the bipartite signal detection software ASAFind (Gruber et al., 2015) utilizing SignalP 3.0 (Bendtsen et al., 2004) and HECTAR (Gschloessl et al., 2008) under default parameters. *Thalassiosira pseudonana* proteins not predicted to be localized to the plastid were further analyzed through sequence homology networks (Kleffmann et al., 2004). Networks were run using EGN (Halary et al., 2013) with an e-value cutoff of 1e-05 and a minimum threshold of 20% similarity over 50% of the shortest peptide, utilizing the following proteome datasets: the plastid proteome of the model plant organism *Arabidopsis thaliana*, cyanobacterial proteomes from 67 distinct strains, and the whole-cell proteome of the oomycete species *Phytophthora sojae* (Tyler et al., 2006).

Resulting networks were filtered for those containing *Thalassiosira pseudonana* proteins and were sorted by the species composition present within the connected components. These networks were sorted into two bins ("Photosynthetic" and "Non-photosynthetic eukaryote") and visualized using Cytoscape (Morris et al., 2012). For proteins with no predicted localization, the detection of potential ungapped sequence motifs for plastid targeting was performed using MEME (Bailey & Elkan, 1994). Motif analysis for these proteins was performed for both full protein sequence and 75 amino acid N-terminals. Further analysis of the 75 amino acid N-terminal sequences was also performed using a sequence similarity network analysis under non-stringent parameters (20% sequence identity across 50% of the shorter peptide). Sequences from resulting network clusters were then submitted for motif analysis using MEME.

In addition, a control group of "ASAFAP" sequences was used to validate sensitivity during network and motif analysis. MEME motif analysis was run under normal mode using a site distribution of zero or one occurrence per sequence and an expected number of motifs of one. Following MEME analysis, resulting motifs were imported into FIMO (Grant et al., 2011), run under default parameters, to be searched against *Arabidopsis thaliana* upstream sequences for functional identification.

RESULTS AND DISCUSSION

Defining a core *Thalassiosira pseudonana* plastid proteome

This study utilized a combination of plastid enrichment and MS-based identification to identify the protein composition of the *Thalassiosira pseudonana* plastid. In addition, by comparing plastid proteomes of *T. pseudonana* cultivated in Fe-limiting and Fe-replete conditions, we investigated how the plastid remodels its proteome in response to altered Fe concentrations. Enrichment of *T. pseudonana* plastids was verified through SDS-PAGE (Figure 2) prior to protein digestion and MS analysis. The large subunit of RuBisCO (rbcL; ~56 kDA) was highly enriched within plastid fractions in comparison to whole-cell fractions, highlighting the enrichment of plastid proteins within our samples.

Across both Fe-replete and Fe-limiting conditions, a total of 929 proteins were identified in the enriched *Thalassiosira pseudonana* plastid proteome in one or both treatments. Within this dataset triplicate technical replicates for each Fe condition, counts for each protein identified exhibited little variation, highlighting the reproducibility of the plastid proteome enrichment (Appendix S1; Table S2 in the Supporting Information). This list includes 73 of the 127 proteins (57%) encoded on the *T. pseudonana* plastid genome (Table S3 in the Supporting Information; Armbrust et al., 2004). The metabolic function of the plastid proteome using Gene Ontology enrichment (Panther; Mi et al., 2013) confirmed the enrichment of plastid proteins in our study, as terms

associated with plastid photosynthetic function, as well as organellar transcription and translation, were highly enriched (Table S4 in the Supporting Information).

Of the total 929 proteins enriched in our plastid proteome, 596 proteins (64%) were confidently identified across enriched plastid proteomes from replicates of diatom cultures grown in Fe-replete and -limiting conditions (≥2 peptides across all replicate Fe-replete and Fe-limited enrichments) and likely represent a constitutive "core" plastid proteome. As these proteins were present across all replicate plastid enrichments (Fe-replete and Fe-limiting), they are highly likely to be plastid-localized or proteins tightly associated with the plastid, including those associated with protein import into the plastid. The *Thalassiosira pseudonana* core proteome in our study contained both plastid- and nucleus-encoded proteins involved in the transcription and translation of plastid-encoded genes, components of plastid ATP synthase, and proteins that serve as components of the plastid light-harvesting complexes, highlighting that the core plastid proteome includes critical translational machinery initiated in the nucleus. Within the core proteome dataset, 135 proteins lack any functional annotation, representing future areas of investigation. Of the 127 proteins encoded on the *T. pseudonana* plastid genome (Armbrust et al., 2004), 55 were present in the core proteome. The remaining 72 proteins include membrane-associated proteins that may be more hydrophobic, making them difficult to isolate, digest, and analyze with MS approaches or not translated under the two Fe conditions investigated here.

As there have been few direct proteome studies of enriched organelles from diatoms, and as plastid enrichment methods differ, it is not surprising that different complements of proteins are identified. A previous analysis of the *Thalassiosira pseudonana* plastid proteome (Schober et al., 2019) identified 217 proteins. The Schober et al. (2019) dataset was particularly enriched in proteins that were membrane associated and embedded, likely resulting from the plastid enrichment and fractionation methods utilized. Of the 217 proteins identified by Schober et al. (2019), we detected 143 (~65%) within our plastid proteome. Analysis of the remaining 74 proteins not located within our dataset shows we may be missing some of the plastid proteins identified by Schober et al. (2019), such as 27 (~33%) that were computationally predicted to be plastid localized by ASAFind (Gruber et al., 2015) and two were encoded on the plastid genome, with the remaining 45 (~60%) not predicted to be plastid localized (Table S2). Of these 74 proteins, 34 are functionally uncharacterized, with the remaining proteins exhibiting membrane-associated function including translocases, or functions involved in vesicle formation and trafficking. The results from Schober et al. (2019) highlight that the stromal fraction may be difficult to maintain in diatom plastid enrichment as they were unable to detect strong enrichment of the stromal marker protein rbcL using Western blots (Schober et al., 2019). This study expands the number of plastid-localized proteins identified in *T. pseudonana*, with enrichment of the stromal fraction as evidenced by enrichment of rbcL (Figure 2). Our study also provides an understanding of plastid protein abundance under changes in Fe status, identifying plastid-localized proteins with presences uniquely identified in only one of the two Fe conditions tested (≥2 peptides detected across all replicate Fe-replete or -limiting enrichments) with 121 proteins unique to the Fe-replete and 212 unique to Fe-limiting conditions. These condition-specific groups included 18 proteins that were significantly differentially abundant (15 in Fe-limiting; three in Fe-replete; Table S2). These differentially expressed proteins, in conjunction with differently abundant proteins in the core proteome, reflect variation in the plastid proteome function in response to changing Fe levels, including changes in light-harvesting complex and Calvin Cycle protein abundance.

Proteome comparison between plastids arising from primary and secondary endosymbiosis

As the plastid proteomes of higher plants that arose from a primary endosymbiotic event have been well characterized, we compared our *Thalassiosira pseudonana* plastid-enriched proteome to that of the plant model species *Arabidopsis thaliana*. The number of plastid-encoded genes in *T. pseudonana* (127) is greater than in the *A. thaliana* plastid genome (80; Table 1), and there are also differences in the complement of nuclear-encoded, plastid-targeted proteins between species (Table S5 in the Supporting Information). Between the *A. thaliana* and *T. pseudonana* plastid proteomes, 139 KEGG functions were shared and represent common plastid functions. These shared KEGG homologies include organelle protein synthesis, photosynthetic function, and carbon fixation. In addition to shared functions, 463 and 308 KEGG terms were unique to *A. thaliana* and *T. pseudonana*, respectively.

One difference was the presence of KEGG terms in the *Thalassiosira pseudonana* plastid proteome associated with endoplasmic reticulum (ER) transport and eukaryotic translation, which can be explained by the evolutionary history of diatoms that resulted in the ER being the outermost membrane of the diatom plastid (Burki et al., 2016). For example, the *T. pseudonana* proteome contains proteins associated with transport into the complex diatom plastid, such as proteins in the Sec61 complex (K10956, K07342; Figure 1), that are required for the import of nuclear-encoded plastid proteins across the outermost membrane of the ER

TABLE 1 Comparison of the number of genes encoded on the plastid proteome and the size of the current confirmed plastid proteomes for *T. pseudonana* and *A. thaliana*.

Organism	Number of plastid-encoded genes	Plastid proteome content	Unique KEGG functions	Shared KEGG functions
T. pseudonana	127	595	308	139
A. thaliana	80	690	463	

Note: Plastid proteome size for *T. pseudonana* listed is the number of plastid- and nuclear-encoded proteins, following protein sorting (Figure 3).

(Gruber et al., 2007; Hempel et al., 2009). Additionally, a total of 417 *T. pseudonana* proteins were unable to be assigned KEGG terms, highlighting the number of putative plastid proteins with unknown function and the potential for differences in plastid function in photosynthetic organisms in lineages with different life histories.

Comparison of proteome composition to computationally predicted plastid localization

Prior research investigating the localization of nuclear-encoded proteins to the diatom plastid identified a highly conserved bi-partite signal peptide of alanine, serine, alanine, phenylalanine, alanine, and proline (ASAFAP) that is responsible for protein targeting to the ER and subsequently into the plastid (Apt et al., 2002; Kroth, 2002). Site-directed mutagenesis of the ASAFAP signal motif has shown that protein import into the plastid is dependent on the highly conserved phenylalanine residue at the +1 position of the cleavage site (Gruber et al., 2007). This motif was used to develop the bioinformatics software, ASAFind (Gruber et al., 2015) that identifies putative plastid-localized proteins in diatoms. Utilizing ASAFind, a total of 1568 nucleus-encoded proteins from the *Thalassiosira pseudonana* genome were computationally predicted to be plastid localized (Gruber et al., 2015).

To compare computationally predicted localization with the protein content of the *Thalassiosira pseudonana* plastid proteome, we analyzed the nucleus-encoded fraction (856 proteins) of the *T. pseudonana* total plastid proteome with localization software (Figure 3). We used both ASAFind and HECTAR, another algorithm designed to predict plastid import of non-plastid-encoded proteins in algae with secondary plastids of the red lineage (Gschloessl et al., 2008). Proteins were assessed as computationally confirmed for plastid import if they were localized to the plastid by either of these programs (Table S2). The combined ASAFind and HECTAR pipeline predicted plastid localization for 175 of the 856 nuclear-encoded proteins in the plastid proteomes (Figure 4), with 538 proteins not computationally predicted to be plastid localized.

The ASAFind and HECTAR analysis of the nucleus-encoded portion of the total *Thalassiosira pseudonana* plastid proteome also predicted localization to other cellular compartments including the ER (116) and mitochondria (27). To ascertain whether proteins predicted to be ER localized by ASAFind and HECTAR might be part of the outer plastid membrane, these proteins were analyzed for transmembrane helix predictions using the trans-membrane helix prediction software TMHMM (Krogh et al., 2001). Fifty-five putative ER membrane-associated proteins with transmembrane domains were predicted and then analyzed using the software suite Mempype (Pierleoni et al., 2011) to bioinformatically sort them into one of three cellular localization bins: cell membrane (17), internal membrane (25), and organellar membrane (13; Table S6 in the Supporting Information). The 13 proteins predicted to be localized to organellar membranes represent proteins likely to be associated with the plastid membrane.

In comparison to the predicted plastid localization of the 175 nucleus-encoded proteins in our proteome, the ASAFind prediction of the entire *Thalassiosira pseudonana* nuclear-encoded proteome includes a total of 1568 proteins that have the potential for plastid localization. Thus, there is a large proportion of *T. pseudonana* proteins predicted to be plastid localized by ASAFind (Gruber et al., 2015; 1393) that are not present in our proteome data. This is likely due to a number of false-positive predictions of localization by computational methods as well as the possibility that some proteins with plastid function are not highly expressed under the growth conditions reported here. Taken together, our data demonstrate that although computational localization prediction is a useful starting tool for proteome analysis, it both potentially overestimates plastid-localized proteins as well as fails to identify them, which may not be an altogether surprising conclusion. We employed further data analysis approaches to see if we could uncover other signals in the proteins that may target them to plastids.

Prediction of protein localization, origin, and function using sequence similarity network analyses

Within our study, a total of 542 proteins in the enriched proteome were not computationally predicted for plastid or ER localization by either ASAFind or HECTAR (Figure 4), suggesting there may be additional

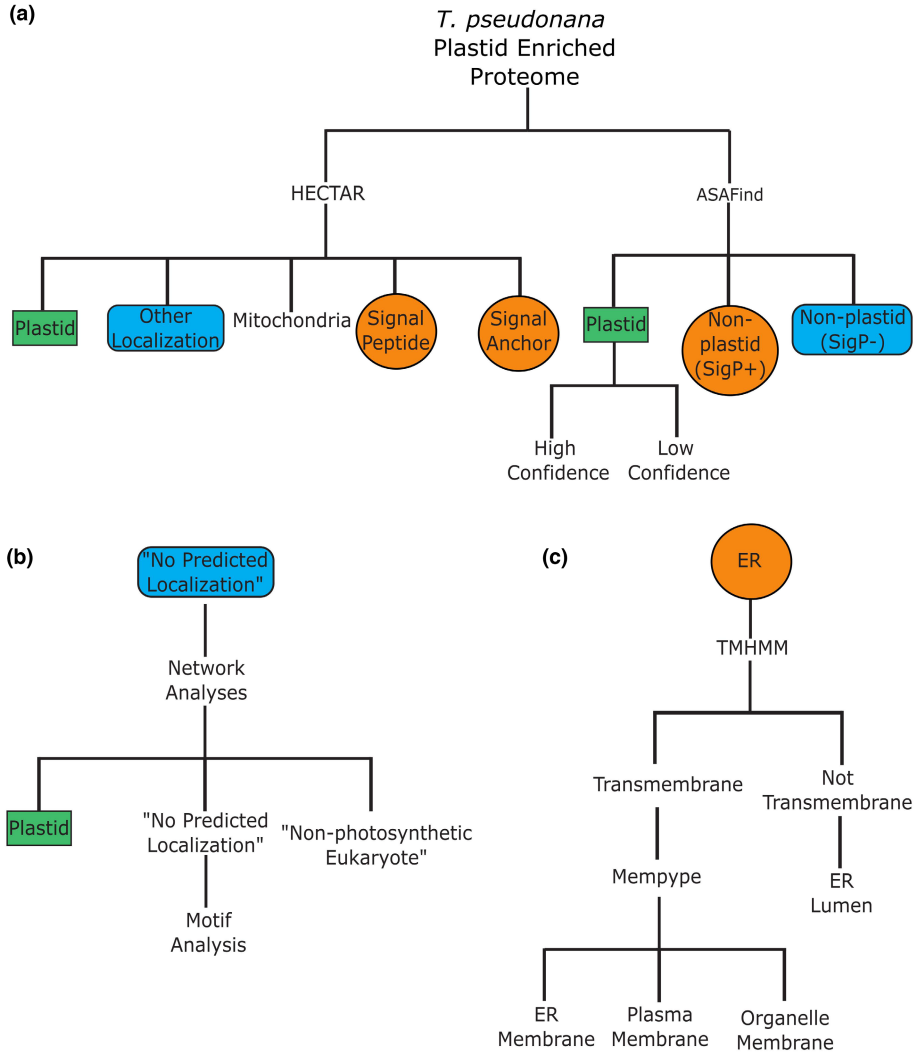

FIGURE 3 Computational pipeline for localization predictions of nuclear-encoded *Thalassiosira pseudonana* plastid proteins. (a) Initial localization prediction of protein localization was performed using ASAFAP and HECTAR, proteins not predicted to the plastid were subjected to additional analyses (b). Proteins with "no predicted localization" following ASAFAP/HECTAR analysis were subjected to homology network analyses to provide further localization. (c) Proteins with predicted signal peptides/anchors were subjected to additional analysis using TMHMM and Mempype to provide further localization. [Color figure can be viewed at wileyonlinelibrary.com]

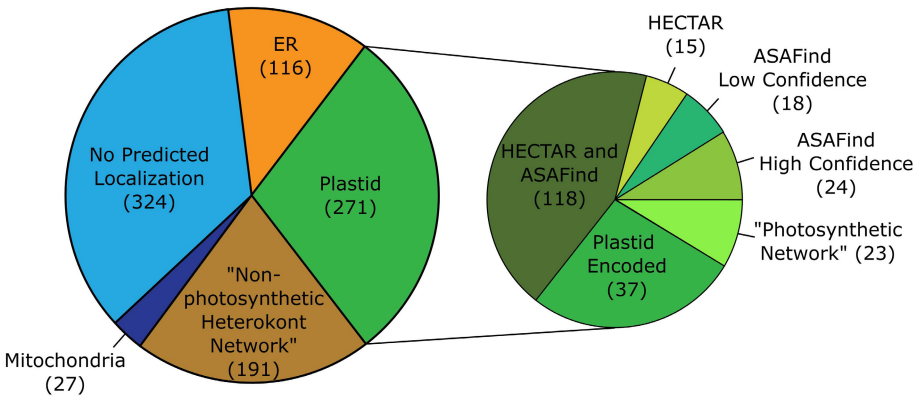

FIGURE 4 Results of computational localization predictions of the 929 proteins from the plastid-enriched proteome of *Thalassiosira pseudonana*. Proteins were analyzed using the plastid localization software, ASAFind, and HECTAR. Proteins with no predicted localizations were subjected to additional analysis using sequence homology networks created using EGN. [Color figure can be viewed at wileyonlinelibrary.com]

organelle-target signals used in diatoms to direct proteins to the plastid. Here we used alternative approaches to interpret the content of our plastid-enriched proteome to decipher whether there may be novel signaling sequences in proteins that were not bioinformatically predicted using current algorithms.

We acknowledge that proteomes derived from plastid enrichment are highly enriched for plastid proteins, but abundant proteins from other cellular components could co-purify with the plastids. To help distinguish plastid proteins from proteins co-purifying with plastids that may function in other parts of the cell, we used a sequence similarity network approach. We compared the group of *Thalassiosira pseudonana* plastid proteins with no predicted localization with three other proteome datasets. To identify proteins of a non-photosynthetic/non-plastid origin, homology comparisons were made to the whole-cell proteome of a non-photosynthetic sister taxa to diatoms, *Phytophthora sojae* (Tyler et al., 2006) in the phylum Oomycota. To identify plastid proteins that potentially arose from the initial primary endosymbiont that gave rise to all plastids, the network analysis incorporated the plastid-enriched proteome from the model plant species *Arabidopsis thaliana* (Kleffmann et al., 2004) along with the combined cyanobacterial proteomes from 107 distinct strains (Table S7; Appendix S2 in the Supporting Information).

Homology network results were used to bin *Thalassiosira pseudonana* proteins with no predicted localization into three categories: those that shared homology with *Phytophthora sojae* proteins, indicated as a "non-photosynthetic eukaryote" bin (191 proteins); those proteins that shared homology between *Arabidopsis* and cyanobacteria proteins, indicated as a "photosynthetic" bin (23 proteins); and those proteins that did not share homology across these proteomes, which those remained designated as "no localization prediction" (324 proteins; Figure 4; Table S2). Within these categorized bins, we did not considered proteins within the "non-photosynthetic eukaryote" bin to have confirmed plastid function. This protein set has high enrichment for GO terms associated with eukaryotic transcription and translation and terms associated with vesicle formation and transport. The enrichment of these terms suggests many of these proteins are associated with the rough ER, the outermost membrane of the complex diatom plastid. Other proteins in this category may be proteins directed to the ER and processed through the eukaryotic secretory pathway. Signal peptide analysis predicts the localization of 116 proteins to the ER within this bin (~60%), where they may be retained, packaged for transport to other cellular compartments, or processed for excretion. Several proteins identified in the "non-photosynthetic eukaryote" group include histones and other eukaryotic DNA binding proteins, as well as mitochondrial electron transport proteins. These proteins are often highly abundant in cells (Huber et al., 2003) and are likely to be protein contaminants in sub-cellular proteomes; these proteins likely co-enriched during the plastid-enriched method.

The group of plastid-enriched proteins with "no localization prediction" may represent diatom plastid proteins with sequences that significantly diverged from other phototrophs. This category includes proteins with canonical plastid function such as light harvesting (B8LE61), thylakoid formation (B8BTT8), and carbonic anhydrases (B8CG97, B8C215) involved in carbon-concentrating mechanisms. The "no localization prediction" category also included proteins that may function exclusively in plastids evolved via secondary endosymbiosis such as Sec61 (B8C4D0) for transport of proteins across the outer plastid membrane and others with homology to proteins involved in ER and Golgi function. Within the "no localization prediction" bin, 46 of the proteins had no predicted function, highlighting the potential for currently undescribed plastid function. However, it is important to note that a number of proteins within this fraction were lacking a starting methionine, indicating a truncated N-terminus, suggesting the need for reassessment of open reading frames (ORFs) within the *Thalassiosira pseudonana* genome and further localization analysis of these proteins.

Assessment of additional localization motifs

The high abundance of plastid proteins lacking the canonical ASAFAP motif suggests the presence of additional plastid targeting mechanisms for nuclear-encoded genes, as has been suggested in other sub-cellular proteome studies in land plants (Kleffmann et al., 2004). To investigate whether there may be plastid-targeting motifs distinct from the ASAFAP motif, the motif discovery program MEME (Bailey et al., 2009) was employed along with protein similarity network analyses to analyze N-terminal peptide homology (Halary et al., 2013). For the 342 proteins in the "no localization prediction" bin, MEME analysis of genome-predicted whole protein sequences revealed no well-conserved motifs. A sequence similarity network was also conducted using the 75 amino acid N-terminal sequences of proteins with "no localization prediction"; only one network of four proteins was formed containing a single motif corresponding to a functional domain of heat shock/chaperone proteins within *Arabidopsis thaliana*. Therefore, no new conserved sequence N-terminal signaling motifs were revealed within the dataset of proteins lacking a predicted origin. The inability to detect novel conserved

N-terminal motifs in these proteins using MEME may indicate the need to reassess predicted ORFs in the *Thalassiosira pseudonana* genome to capture these N-terminal motifs, or it may indicate that some diatom plastid targeting motifs may be degenerate gapped motifs, dependent on charge-based secondary structures, such as is observed in mitochondrial protein import (Claros & Vincens, 1996; von Heijne, 1986).

The *Thalassiosira pseudonana* plastid proteins comprising those in the bioinformatically filtered photosynthetic bin and the bin of proteins with "no localization prediction" (Figure 4) were used to curate a defined set of nucleus-encoded plastid proteins. These proteins represent a potential expansion of known plastid-targeted proteins, and future functional analyses of these targets may provide insight into additional protein import mechanisms in the complex secondary plastid. The large number of proteins lacking functional homology (228) within this dataset represents potentially new functions that have yet to be described in diatom plastid biogenesis and metabolism.

Fe-driven remodeling of the plastid-enriched proteome

Under Fe-limitation, we observed statistically significant changes in abundance of 135 proteins (~22% of the proteome) within our plastid-enriched proteome. Of those 135 proteins responding in the Fe-limiting condition, 77 increased in abundance and 58 decreased in abundance relative to Fe-replete conditions. Proteins present within this Fe-regulated fraction represented multiple plastid functions, including carbon fixation and light stress responses, exhibiting strong plastid response to changes in environmental Fe levels, and included 73 Fe-sensitive proteins with uncharacterized function.

Components of the Calvin cycle are downregulated in response to iron stress

Carbon fixation through the Calvin cycle represents one of the main metabolic functions of the photosynthetic plastid. Carbon fixation is driven by ATP and NADPH, two byproducts of the photosynthetic electron transport chain where Fe is a critical co-factor (Raven, 1990; Raven et al., 1999; van Oijen et al., 2004). Therefore, understanding the impacts of Fe limitation on the electron transport chain and downstream carbon fixation is crucial for modeling Fe-driven environmental impacts. The proteins phosphoglycerate-kinase (B5YN92), triose-phosphate isomerase (B8C8U5), fructose-bisphosphate aldolase (B8BXS1), and transketolase (B8BTR4) are found within the reductive and regenerative phases of the Calvin Cycle. The observed decrease in abundance of these proteins under Fe-limitation in our study suggests reduced carbon fixation within the plastid, in response to decreased photosynthetic efficiency under Fe-limiting conditions (Table 2).

In contrast to prior whole-cell proteome studies of Fe-limited *Thalassiosira pseudonana* (Nunn et al., 2013), in which transketolase, fructose-bisphosphate aldolase, and phosphoglycerate kinase exhibited increases in abundance, we observed a decrease in abundance within the Fe-limited plastid-enriched proteome. The contrasting results may stem from the participation of these proteins in multiple metabolic pathways that are carried out in different cellular components (i.e., cytoplasm-directed glycolysis). Using sub-cellular proteomics, we have been able to discern alterations in the abundance of the plastid-localized copies of these proteins and their role in the Calvin cycle under Fe limitation. Another point of contrast with whole-cell *T. pseudonana* proteomes is that we did not see differences in RuBisCO large subunit (A0T0N6) expression, whereas it was present at increased abundances under

TABLE 2 Inventory of *Thalassiosira pseudonana* encoded Calvin Cycle proteins within +/Fe-plastid proteomes and their regulation under Fe-limited condition.

Name	Uniprot ID	Change in abundance	Log_2 fold change
RuBisCO large subunit	A0T0N6	No change	−0.234
RuBisCO small subunit	A0T0N5	No change	−0.475
Phosphoglycerate-kinase	B5YN92	Decreased	−0.554
Glyceraldehyde-3-phosphate dehydrogenase	B8BQU2	No change	0.322
Triose-phosphate isomerase	B8C8U5	Decreased	−2.216
Fructose 1,6-bisphosphatase	Not detected	–	–
Fructose-bisphosphate aldolase	B8BXS1	Decreased	−1.849
Transketolase	B8BTR4	Decreased	−1.248
Ribose-5-phosphate isomerase	Not detected	–	–
Phosphopentose epimerase	Not detected	–	–
Phosphoribulokinase	Not detected	–	–

Fe-limiting conditions in whole-cell proteomes (Nunn et al., 2013).

Iron limitation alters light harvesting protein abundance

As mentioned earlier, Fe limitation stresses photosynthetic plastid function, as Fe is an important co-factor in the photosynthetic electron transport chain (Raven, 1990; Raven et al., 1999; van Oijen et al., 2004). Additionally, as a result of excess light excitation, Fe limitation can cause photoinhibition (Zhu et al., 2010) and photoinactivation of photosystem subunits, including protein D1 (A0T0W2) of Photosystem II (Li et al., 2016). In comparison to green algae, diatoms contain expanded families of fucoxanthin chlorophyll *a*/*c*-binding proteins (FCPs), including the Light-Harvesting Complex X (LHCX) clade of FCPs, with the number of isoforms differing between diatom species (Taddei et al., 2016). Here, three of the five LHCX proteins encoded within the *Thalassiosira pseudonana* genome were detected in plastid fractions, and LHCX6 was present with significantly increased abundance under Fe limitation (Table 3).

In response to conditions like Fe limitation that induce photoinhibition (Li et al., 2016; Zhu et al., 2010), diatoms alter their ratios of antenna complexes including fucoxanthin chlorophyll *a*/*c* binding proteins, such as LHCX proteins, to dissipate excess light through non-photochemical quenching (NPQ; Giovagnetti et al., 2021; Lepetit et al., 2016). LHCX1 has been suggested to serve as the primary means of NPQ, as it is constitutively expressed in response to dynamic light (Giovagnetti et al., 2021; Lepetit et al., 2016). We likewise observed a constitutively high count for LHCX1 (B8CGG0) under both Fe-replete and Fe-limiting conditions and constitutive abundance of LHCX5 (B8BSG2) levels across both Fe conditions, albeit at levels lower than observed for LHCX1. This lower constitutive abundance suggests LHCX5 may play an accessory role in general NPQ in conjunction with LHCX1.

TABLE 3 Inventory of *Thalassiosira pseudonana* encoded Lhcx proteins within Fe+/Fe-plastid proteomes and their regulation under Fe-limited condition.

Name	Uniprot ID	Change in abundance	Log_2 fold change
LHCX1	B8CGG0	No change	−0.038
LHCX2	B8CGG2	–	–
LHCX4	B8C364	–	–
LHCX5	B8BSG2	No change	−0.363
LHCX6	B8CGG1	Increased	1.945
LHCX6_1	B5YLU3	No change	0.020

Diatom species also differ in their composition of LHCX proteins and in their patterns of abundance. Under Fe-limitation, LHCX6 (B8CGG1) was uniquely detected, resulting in a reported significant increase in abundance. In *Thalassiosira pseudonana* this protein may present an alternative means of Fe-limitation-induced NPQ response, as it has been shown to be induced under high-light conditions within the polar diatom species *Chaetoceros neogracile* (Park et al., 2010). In *Phaeodactylum tricornutum*, Fe-limitation increased LHCX2 transcripts, with a slight induction of the other LHCX isoforms (Taddei et al., 2016). In contrast, LHCX2 in *T. pseudonana* was not detected under either Fe-replete or Fe-limiting conditions, suggesting differing responses between the low Fe-tolerant *P. tricornutum* and the higher-Fe requiring *T. pseudonana*. We did not observe peptides for LHCX4 within the plastid-enriched proteome. This isoform has been shown to be associated with dark-acclimated cells and is highly down-regulated under even low-light conditions (Lepetit et al., 2013; Nymark et al., 2013). As plastids enriched in this experiment were not from dark-acclimated cells, this is consistent with the absence of LHCX4 within the dataset.

Iron limitation may increase recycling of the PSII protein D1

Diatom plastid genomes maintain genes of prokaryotic origin, including transcriptional and translational machinery, distinct from the gene expression machinery of the eukaryotic nucleus (Hadariová et al., 2017). We observed statistically significant higher abundances under Fe-limitation of plastid-encoded proteins involved in transcription (A0T0Q9, A0T0R0, A0T0Z4) and translation (A0T0Q8, A0T0X9, A0T0Y3, A0T0Y5, A0T0Z5), suggesting the plastid is compensating for photosynthetic protein loss by synthesizing additional plastid-encoded proteins. One such protein that may be actively resynthesized under Fe limitation is the PSII protein D1 (A0T0W2). D1 can be inactivated due to excess light excitation (Li et al., 2016) requiring resynthesis. To preserve photosynthetic processes, diatoms cycle inactivated subunits (Figure 5) through a pool of excess active subunits and degradation of inactive D1 subunits with specific proteases, including the zinc metalloprotease FtsH (A0T0S3; Kato & Sakamoto, 2009; Li et al., 2016). No significant change in abundance of D1 was detected (log_2 fold change = 0.322; A0T0W2) under Fe-limiting conditions. It should be noted that FtsH exhibits constitutive abundance across Fe levels (log_2 fold change = 0.068), suggesting that Fe-replete levels of FtsH is capable of meeting rates of photoinactivation of D1 under low Fe conditions. Within this dataset, we observed no statistically significant increase or decrease in any other plastid-encoded proteins.

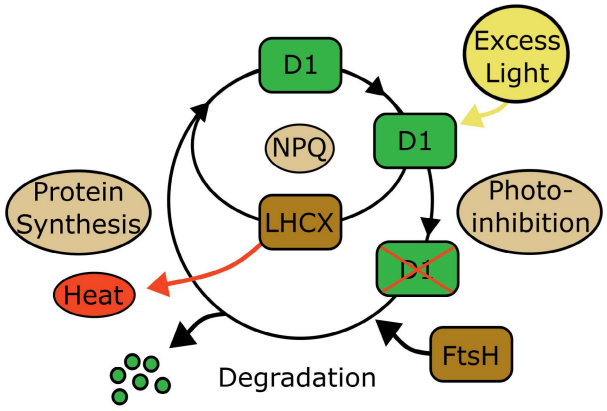

FIGURE 5 Recycling of Photosystem II subunits during excess light stimulation. Upon excess light excitation PSII D1 can become photoinactivated and subsequently degraded by the metalloprotease FtsH. Degraded D1 is then replaced by subunits synthesized within the plastid. Photo-inhibition can be prevented through the process of non-photochemical quenching (NPQ), during which excess light energy is dissipated as heat by light-harvesting complex proteins (LHC). This dissipation of excess light energy by NPQ, prevents the photo-inhibition of D1. [Color figure can be viewed at wileyonlinelibrary.com]

CONCLUSIONS AND PERSPECTIVE

Sub-cellular enrichment of organelles represents an important approach for understanding diatom protein localization and changes in their abundance in specific organelles. While whole-cell proteomic methods allow for a broad overview of changes in protein abundance, they often suffer from a lack of sensitivity due to detection limits of lower abundance organellar proteins. This issue is overcome by using organelle enrichments prior to MS-based analyses because these enrichments increase detection sensitivity and reduce the complexity of proteins to be analyzed, facilitating the identification of organelle-specific proteins and direct observations of the compartmentalization of different metabolic processes. In addition, proteomics and sub-cellular proteomics allows for simultaneous quantification of plastid-encoded and plastid-targeted proteins, analyses that are not possible with transcriptomic data using standard poly(A) selection for mRNA enrichment, as organelle-encoded genes lack poly(A) tails.

Plastid proteome sequencing in the diatom *Thalassiosira pseudonana*, in this study and in Schober et al. (2019), has confirmed the identification of the plastid-localized proteome in this model species and facilitates comparison to predictive studies of plastid protein targeting. The number of plastid proteins identified in this study that lack a canonical N-terminal ASAFAP (538) motif suggests the presence of yet uncharacterized, ASAFAP motif-independent plastid targeting pathways for nucleus-encoded proteins. The discrepancy to the high number of computationally predicted plastid-targeted proteins (Gruber et al., 2015) that are not detected in the plastid-enriched proteome suggests that the full inventory of proteins localized to the plastid may be controlled by physiology-specific factors such as differences in nutrient supply or growth state. Our direct detection of a unique suite of plastid proteins in the Fe-limiting cultures supports this evidence-based theory. Detailed examinations of the protein-level response detected in divergent Fe conditions revealed changes in the abundance of light-harvesting complex proteins, Calvin cycle proteins (specifically the regenerative phase), and PSII D1 recycling rates.

The identification of a portion of proteins in our *Thalassiosira pseudonana* proteome that are functionally uncharacterized emphasizes the continued need for functional characterization studies in diatoms. The comparison of chloroplast proteomes under Fe sufficiency and limitation helps define *T. pseudonana* plastid constituents particularly sensitive to Fe and highlights that our current understanding of diatom protein function may underestimate how Fe bioavailability may impact plastid function. To assess differences in the Fe-limiting response of the plastid-enriched proteome as a potential factor in differences of Fe requirements between diatom species, future analyses of closely related *Thalassiosira* species with low Fe requirements may provide insight into different adaptive responses and additional Fe-level sensitive proteins localized to within the plastid. In addition, a comparison of plastid-enriched proteome metabolic function from distinct photosynthetic lineages will provide insight into the evolution of plastid-localized metabolic functions.

AUTHOR CONTRIBUTIONS

Kristofer M. Gomes: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); visualization (equal); writing – original draft (lead); writing – review and editing (equal). **P. Dreux Chappell:** Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal); visualization (equal); writing – review and editing (equal). **Brook L. Nunn:** Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); validation (equal); visualization (equal); writing – review and editing (equal). **Bethany D Jenkins:** Conceptualization (equal); funding acquisition (lead); investigation (equal); methodology (equal); supervision (lead); writing – review and editing (equal).

ORCID

Kristofer M. Gomes https://orcid.org/0000-0002-7645-6018
Brook L. Nunn https://orcid.org/0000-0002-7361-4359
P. Dreux Chappell https://orcid.org/0000-0001-5212-6228
Bethany D. Jenkins https://orcid.org/0000-0003-2319-3098

REFERENCES

Allen, A. E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G. F., Fernie, A. S., & Bowler, C. (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proceedings of the National Academy of Science*, *105*(30), 10438–10443.

Allen, J. F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. *Proceedings of the National Academy of Science*, *112*(33), 10231–10238.

Apt, K. E., Zaslavkaia, L., Lippmeier, J. C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A. R., & Kroth, P. G. (2002). In vivo characterization of diatom multipartite plastid targeting signals. *Journal of Cell Science*, *115*(221), 4061–4069.

Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Current Biology*, *25*(19), R911–R921.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., … Rokhsar, D. S. (2004). The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science*, *306*(5693), 79–86.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*, 202–208.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, *2*, 28–36.

Barkan, A. (1993). Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. *Plant Cell*, *5*(4), 389–402.

Bendtsen, J. D., Nielsen, H., von Heijne, G., & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology*, *340*(4), 783–795.

Bhattacharya, D., Archibald, J. M., Weber, A. P. M., & Reyes-Prieto, A. (2007). How do endosymbionts become organelles? Understanding early events in plastid evolution. *BioEssays*, *29*(12), 1239–1246.

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B. Q., Radaykina, L. V., Smirnov, A., Mylnikov, A. P., & Keeling, P. (2016). Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proceedings of the Royal Society B Biological Sciences*, *283*(1823), 20152802.

Chappell, P. D., & Jenkins, B. D. (2017). Diatom chloroplast isolation steps version 3. https://www.protocols.io/

Choi, H., Fermin, D., & Nesvizhskii, A. I. (2008). Significance analysis of spectral count data in label-free shotgun proteomics. *Molecular and Cellular Proteomics*, *7*(12), 2373–2385.

Claros, M. G., & Vincens, P. (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry*, *241*(3), 779–786.

Cohen, N. R., Gong, W., Moran, D. M., McIlvin, M. R., Saito, M. A., & Marchetti, A. (2018). Transcriptomic and proteomic responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation. *Environmental Microbiology*, *20*(8), 3109–3126.

Elias, J. E., & Gygi, S. P. (2010). Target-decoy search strategy for mass spectrometry-based proteomics. *Methods in Molecular Biology*, *604*, 55–71.

Eng, J. K., Jahan, T. A., & Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics*, *13*(1), 22–24.

Fermin, D., Basrur, V., Yocum, A. K., & Nesvizhskii, A. I. (2011). Abacus: A computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics*, *11*(7), 1340–1345.

Friso, G., Giacomelli, L., Ytterberg, A., & Peltier, J. (2004). In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana*. *Plant Cell*, *16*(2), 478–949.

Gao, X., Bowler, C., & Kazamia, E. (2021). Iron metabolism strategies in diatoms. *Journal of Experimental Botany*, *72*(6), 2165–2180.

Geider, R. J., & la Roche, J. (1994). The role of iron in phytoplankton photosynthesis, and the potential for iron-limitation of primary productivity in the sea. *Photosynthesis Research*, *39*, 275–301.

Giovagnetti, V., Jaubert, M., Shukla, M. K., Ungerer, P., Bouly, J. P., Falciatore, A., & Ruban, A. V. (2021). Biochemical and molecular properties of LHCX1, the essential regulator of dynamic photoprotection in diatoms. *Plant Physiology*, *188*(1), 509–525.

Graff van Creveld, S., Coesel, S., Blaskowski, S., Groussman, R., Schatz, M., & Armbrust, V. (2023). Divergent functions of two clades of flavodoxin in diatoms mitigate oxidative stress and iron limitation. *eLife*, *12*, e84392.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics*, *27*(7), 1017–1018.

Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., & Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant Journal*, *81*(3), 519–528.

Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U. G., & Kroth, P. G. (2007). Protein targeting into complex diatom plastids: Functional characterisation of a specific targeting motif. *Plant Molecular Biology*, *64*(5), 519–530.

Gschloessl, B., Guermeur, Y., & Cock, J. M. (2008). HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*, *9*, 393.

Guillard, R. R. L., & Ryther, J. H. (1962). Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt, and *Detonula confervacea* (cleve) gran. *Canadian Journal of Microbiology*, *8*, 229–239.

Hadariová, L., Vesteg, M., Hampl, V., & Krajčovič, J. (2017). Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Current Genetics*, *64*(2), 365–387.

Halary, S., McInerney, J. O., Lopez, P., & Bapteste, E. (2013). EGN: A wizard for construction of gene and genome similarity networks. *BMC Evolutionary Biology*, *13*, 146.

Hempel, F., Bullmann, L., Lau, J., Zauner, S., & Maier, U. G. (2009). ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Molecular Biology and Evolution*, *26*(8), 1781–1790.

Huber, L. A., Pfaller, K., & Vietor, I. (2003). Organelle proteomics: Implications for subcellular fractionation in proteomics. *Circulation Research*, *92*(9), 962–968.

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2018). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314.

Huesgen, P. F., Alami, M., Lange, P. F., Foster, L. J., Schröder, W. P., Overall, C. M., & Green, B. R. (2013). Proteomic amino-termini profiling reveals targeting information for protein import into complex plastids. *PLoS ONE*, *8*(9), e74483.

Hutchins, D. A., & Boyd, P. W. (2016). Marine phytoplankton and the changing ocean iron cycle. *Nature Climate Change*, *6*, 1072–1079.

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of

genome and metagenome sequences. *Journal of Molecular Biology*, *428*(4), 726–731.

Kato, Y., & Sakamoto, W. (2009). Protein quality control in chloroplasts: A current model of D1 protein degradation in the photosystem II repair cycle. *The Journal of Biochemistry*, *146*(4), 463–469.

Keller, A., & Shteynberg, D. (2011). Software pipeline and data analysis for MS/MS proteomics: The trans-proteomic pipeline. In C. H. Wu & C. Chen (Eds.), *Bioinformatics for comparative proteomics* (pp. 169–189). Springer.

Kleffmann, T., Russenberger, D., Von Zychlinski, A., Christopher, W., Gruissem, W., & Baginsky, S. (2004). The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Current Biology*, *14*(5), 354–362.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580.

Kroth, P. G. (2002). Protein transport into secondary plastids and the evolution of primary and secondary plastids. *International Review of Cytology*, *221*, 191–255.

Lepetit, B., Gélin, G., Lepetit, M., Sturm, S., Vugrinec, S., Rogato, A., Kroth, P. G., Falciatore, A., & Lavaud, J. (2016). The diatom *Phaeodactylum tricornutum* adjusts nonphotochemical fluorescence quenching capacity in response to dynamic light via fine-tuned Lhcx and xanthophyll cycle pigment synthesis. *New Phytologist.*, *214*(1), 205–218.

Lepetit, B., Sturm, S., Rogato, A., Gruber, A., Sachse, M., Falciatore, A., Kroth, P. G., & Lavaud, J. (2013). High light acclimation in the secondary plastids containing diatom *Phaeodactylum tricornutum* is triggered by the redox state of the plastoquinone pool. *Plant Physiology*, *161*(2), 853–865.

Li, G., Woroch, A. D., Donaher, N. A., Cockshutt, A. M., & Campbell, D. A. (2016). A hard day's night: Diatoms continue recycling photosystem II in the dark. *Frontiers in Marine Science*, *3*, 218.

Martin, J., Fitzwater, S., & Gordon, R. M. (1990). Iron deficiency limits phytoplankton growth in Antarctic waters. *Global Biogeochemical Cycles*, *4*(1), 5–12.

Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, *41*, 377–386.

Milligan, A. J., & Harrison, P. J. (2000). Effects of non-steady-state iron limitation on nitrogen assimilation in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). *Journal of Phycology*, *36*(1), 78–86.

Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., Durkin, C., BonDurant, S. S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E. L., Cerrina, F., Sussman, M. R., & Armbrust, E. V. (2008). Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Science*, *105*(5), 1579–1584.

Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C., & Misumi, K. (2013). Marine ecosystem dynamics and iogeochemical cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 scenarios. *Journal of Climate*, *26*(23), 9291–9312.

Moreno, C. M., Gong, W., Cohen, N. R., DeLong, K., & Marchetti, A. (2020). Interactive effects of iron and light limitation on the molecular physiology of the Southern Ocean diatom *Fragilariopsis kerguelensis*. *Limnology and Oceanography*, *65*(7), 1511–1531.

Moreno, C. M., Lin, Y., Davies, S., Monbureau, E., Cassar, N., & Marchetti, A. (2017). Examination of gene repertoires and physiological responses to iron and light limitation in Southern Ocean diatoms. *Polar Biology*, *41*, 679–696.

Morris, J. S., Ph, D., Kuchinsky, A., Pico, A., & Institutes, G. (2012). Analysis and visualization of biological networks with Cytoscape. [UCSF Workshop].

Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., Thomas, Y., Botebol, H., Bouget, F. Y., McQuaid, J. B., Tirichine, L., Allen, A. E., Lesuisse, E., & Bowler, C. (2015). A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. *Current Biology*, *25*(3), 364–371.

Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A., & Queguiner, B. (1995). 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), 359–372.

Nunn, B. L., Faux, J. F., Hippmann, A. A., Maldonado, M. T., Harvey, H. R., Goodlett, D. R., Boyd, P. W., & Strzepek, R. F. (2013). Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLoS O*, *8*(10), e75653.

Nymark, M., Valle, K. C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., Bones, A. M., & Brembu, T. (2013). Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. *PLoS ONE*, *8*(3), e58722.

Park, S., Jung, G., Hwang, Y. S., & Jin, E. (2010). Dynamic response of the transcriptome of a psychrophilic diatom, *Chaetoceros neogracile*, to high irradiance. *Planta*, *231*(2), 349–360.

Pierleoni, A., Indio, V., Savojardo, C., Fariselli, P., Martelli, P. L., & Casadio, R. (2011). MemPype: A pipeline for the annotation of eukaryotic membrane proteins. *Nucleic Acids Research*, *39*, 375–380.

Raven, J. A. (1990). Predictions of Mn and Fe use efficiencies of phototrophic growth as a function of light availability for growth and of C assimilation pathway. *New Phytologist*, *116*(1), 1–18.

Raven, J. A., Evans, M. C. W., & Korb, R. E. (1999). The role of trace metals in photosynthetic electron transport in O_2-evolving organisms. *Photosynthesis Research*, *60*, 111–150.

Schober, A. F., Río Bártulos, C., Bischoff, A., Lepetit, B., Gruber, A., & Kroth, P. G. (2019). Organelle studies and proteome analyses of mitochondria and plastids fractions from the diatom *Thalassiosira pseudonana*. *Plant and Cell Physiology*, *60*(8), 1811–1828.

Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., & Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nature Communications*, *5*, 5764.

Stork, S., Moog, D., Przyborski, J. M., Wilhelmi, I., Zauner, S., & Maier, U. G. (2012). Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic Cell*, *11*(12), 1472–1481.

Sunda, W. G., & Huntsman, S. A. (1995). Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Marine Chemistry*, *50*(1), 189–206.

Taddei, L., Stella, G. R., Rogato, A., Bailleul, B., Fortunato, A. E., Annunziata, R., Sanges, R., Thaler, M., Lepetit, B., Lavaud, J., Jaubert, M., Finazzi, G., Bouly, J. P., & Falciatore, A. (2016). Multisignal control of expression of the LHCX protein family in the marine diatom *Phaeodactylum tricornutum*. *Journal of Experimental Botany*, *67*(13), 3939–3951.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M., Dorrance, A. E., Dou, D., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., … Boore, J. L. (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science*, *313*(5791), 1261–1266.

van Oijen, T., van Leeuwe, M. A., Gieskes, W. W. C., & de Baar, H. J. W. (2004). Effects of iron limitation on photosynthesis and

carbohydrate metabolism in the Antarctic diatom *Chaetoceros brevis* (Bacillariophyceae). *European Journal of Phycology*, *39*(2), 161–171.
von Heijne, G. (1986). Mitochondrial targeting sequences may form amphiphilic helices. *The EMBO Journal*, *5*(6), 1335–1342.
Wang, J., Yu, Q., Xiong, H., Wang, J., Chen, S., Yang, Z., & Dai, S. (2016). Proteomic insight into the response of Arabidopsis chloroplasts to darkness. *PLoS ONE*, *11*(5), e0154235.
Whitney, L. P., Lins, J. J., Hughes, M. P., Wells, M. L., Chappell, P. D., & Jenkins, B. D. (2011). Characterization of putative iron responsive genes as species-specific indicators of iron stress in *Thalassiosiroid* diatoms. *Frontiers in Microbiology*, *2*, 234.
Zhu, S. H., Guo, J., Maldonado, M. T., & Green, B. R. (2010). Effects of iron and copper deficiency on the expression of members of the light-harvesting family in the diatom *Thalassiosira pseudonana* (bacillariophyceae). *Journal of Phycology*, *46*(5), 974–981.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. UniProt sequences of proteins identified within the *Thalassiosira pseudonana* plastid enriched proteome.

Appendix S2. Protein sequences of the cyanobacterial reference database used in localization network analyses.

Table S1. Growth curves of *Thalassiosira pseudonana* under Fe-replete and Fe-limiting conditions during culture acclimation period proceeding harvesting for plastid enrichment. Specific growth rates for each culturing condition were calculated for the period proceeding harvesting.

Table S2. The 929 proteins identified within the *Thalassiosira pseudonana* plastid-enriched proteome. UniProt IDs, counts for each enrichment replicate, log2fold change, FDR, ASAFind/HECTAR localization predictions, network analysis categorization, and KEGG/GO terms are provided for each protein. A comparison of *T. pseudonana* plastid proteome content between this study and Schober et. al (2019) has also been provided.

Table S3. Comparison of plastid genome encoded proteins and their presence within the *Thalassiosira pseudonana* plastid enriched proteome. Protein complexes have been colored by the number of proteins identified within the study (green: all subunits identified, yellow: one or more subunit identified, red: no subunits identified.)

Table S4. Results of PANTHER Overrepresentation Test on the *Thalassiosira pseudonana* plastid enriched proteome.

Table S5. Comparison of KEGG terms for proteins identified within the *Thalassiosira pseudonana* plastid enriched proteome and the *Arabidopsis thaliana* plastid proteome (Kleffmann et al., 2004).

Table S6. Mempype Localization Prediction results for putative ER membrane-associated proteins with transmembrane domains.

Table S7. Cyanobacterial species with proteomes present on UniProt that were used in the generation of the cyanobacterial reference database (Appendix S2) used in localization network analyses.

How to cite this article: Gomes, K. M., Nunn, B. L., Chappell, P. D., & Jenkins, B. D. (2023). Subcellular proteomics for determining iron-limited remodeling of plastids in the model diatom *Thalassiosira pseudonana* (Bacillariophyta). *Journal of Phycology*, *59*, 1085–1099. https://doi.org/10.1111/jpy.13379