DiatOmicBase, a gene-centered platform to mine functional omics data across diatom genomes

Emilie Villar, Nathanaël Zweig, Pierre Vincens, Helena Cruz de Carvalho, Carole Duchene, Shun Liu, Raphael Monteil, Richard G Dorrell, Michele Fabris, Klaas Vandepoele, et al.

DiatOmicBase, a gene-centered platform to mine functional omics data across diatom genomes

Emilie Villar[1,2], Nathanaël Zweig[1], Pierre Vincens[1], Helena Cruz de Carvalho[1,3], Carole Duchene[4,§], Shun Liu[1,#], Raphael Monteil[4], Richard G. Dorrell[5], Michele Fabris[6], Klaas Vandepoele[7,8,9], Chris Bowler[1,*] & Angela Falciatore[4,5,*]

1 - Institut de Biologie de l'École Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
2 - EV Consulting, Marseille, France
3 - Faculté des Sciences et Technologie, Université Paris Est-Créteil (UPEC), 94000 Créteil, France
4 - Institut de Biologie Physico-Chimique, Laboratory of Chloroplast Biology and Light Sensing in Microalgae, UMR7141 Centre National de la Recherche Scientifique (CNRS), Sorbonne Université, 75005 Paris, France
5- CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative - UMR 7238, Sorbonne Université, 4 place Jussieu, 75005 Paris, France
6 - SDU Biotechnology, Department of Green Technology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
7 - Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium
8 - VIB-UGent Center for Plant Systems Biology, Technologiepark 71, 9052 Ghent, Belgium
9 - VIB Center for AI & Computational Biology, VIB, Ghent, Belgium

§current address: Department of Algal Development and Evolution, Max Planck Institute for Biology, 72076 Tuebingen, Germany
#current address: Guangzhou Marine Geological Survey, Guangzhou, China

*Corresponding authors: angela.falciatore@ibpc.fr, cbowler@biologie.ens.fr

Abstract :

Diatoms are prominent microalgae found in all aquatic environments. Over the last 20 years, thanks to the availability of genomic and genetic resources, diatom species such as *Phaeodactylum tricornutum* have emerged as valuable experimental model systems for exploring topics ranging from evolution to cell biology, (eco)physiology and biotechnology. Since the first genome sequencing in 2008, numerous genome-enabled datasets have been generated, based on RNA-Seq and proteomics, epigenomes, and ecotype variant analysis. Unfortunately, these resources, generated by various laboratories, are often in disparate formats and challenging to access and analyze. Here we present DiatOmicBase, a genome portal gathering comprehensive omics resources from *P. tricornutum* and two other diatoms to facilitate the exploration of dispersed public datasets and the design of new experiments based on the prior-art.

DiatOmicBase provides gene annotations, transcriptomic profiles and a genome browser with ecotype variants, histone and methylation marks, transposable elements, non-coding RNAs, and read densities from RNA-Seq experiments. We developed a semi-automatically updated transcriptomic module to explore both publicly available RNA-Seq experiments and users' private datasets. Using gene-level expression data, users can perform exploratory data analysis, differential expression, pathway analysis, biclustering, and co-expression network analysis. Users can create heatmaps to visualize precomputed comparisons for selected gene subsets. Automatic access to other bioinformatic resources and tools for diatom comparative and functional genomics is also provided. Focusing on the resources currently centralized for *P. tricornutum*, we showcase several examples of how DiatOmicBase strengthens molecular research on diatoms, making these organisms accessible to a broad research community.

Significance statement :

In recent years, diatoms have become the subject of increasing interest because of their ecological importance and their biotechnological potential for natural products such as pigments and polyunsaturated fatty acids. Here, we present an interactive web-based server that integrates public diatom 'omics data (genomics, transcriptomics, epigenomics, proteomics, sequence variants) to connect individual diatom genes to broader-scale functional processes.

Keywords :

Diatoms, *Phaeodactylum tricornutum*, genome portal, genome browser, gene models, ecotype variants, histone marks, non-coding RNAs, protein domains, RNA-Seq datasets

Introduction:

Diatoms are unicellular algae that play a major ecological role by contributing up to 20% of carbon fixation in aquatic ecosystems (Tréguer et al., 2021). The group encompasses up to 100,000 species (Alverson et al., 2007; Malviya et al., 2016) with a large diversity of morphologies, sizes and life histories. They are classified into two morphogroups according to their shape: centric diatoms are radially symmetric while pennates are bilaterally symmetrical. Their main common characteristic is the silica cell wall, or “frustule”, that make diatoms key players in the oceanic silica biogeochemical cycle (Tréguer et al., 2021) in addition to the carbon cycle. Widely distributed in any aquatic or humid environments, diatoms are particularly abundant in nutrient-rich coastal ecosystems as well as at high latitudes (Malviya et al., 2016). The diverse habitats in which they occur reflect their extreme adaptive capacities: they can be planktonic and benthic, and can be found even as epiphytes in terrestrial forests, in soil or associated with sea ice (Vanormelingen et al., 2009, Singer et al., 2021).

Besides their ecological significance, diatoms have recently emerged as interesting novel experimental systems to explore still largely uncharacterized features of phytoplankton biology (Falciatore and Mock, 2022). Two species are widely used to decipher diatom biology and molecular functioning: *Thalassiosira pseudonana* (a centric diatom) and *Phaeodactylum tricornutum* (a pennate diatom). They both present significant advantages: they are easily cultivated in the laboratory with rapid growth rates, their genetic transformation is mastered and they have small genomes (32.1 and 27.4 Mb, respectively) encoding around 12,000 genes (Armbrust et al., 2004; Bowler et al., 2008). In recent years, a growing number of different genomic, genetic, physiological and metabolic information and resources have been generated for these two species, making them suitable models to study diatom gene functions and metabolic pathways (Broddrick et al., 2019; Falciatore et al., 2020; Poulsen and Kröger, 2023). In parallel, additional diatom species have also been proposed as models to understand specific eco-physiological adaptations (*e.g.*, *Fragilariopsis cylindrus* for polar habitats (Mock et al., 2017), *Thalassiosira oceanica* for open-oceans (Lommer et al., 2012), and *Seminavis robusta (Osuna-Cruz et al., 2020)* for benthic

106 environments) or to address specific features of diatom life cycles (e.g., *Pseudo-nitzschia*
107 *multistriata* for studying diatom sex (Ferrante et al., 2023)).
108 As of March 2024, 117 complete genomes of pennate and centric diatoms have been
109 deposited in Genbank, and an ongoing project has announced the sequencing of 100 new
110 diatom genomes (JGI initiative). The Marine Microbial Eukaryotic Transcriptome
111 Sequencing Project (MMETSP) has additionally provided 92 different transcriptomes from
112 diverse diatom species (Keeling et al., 2014) which has been recently completed with 6 other
113 transcriptomes (Dorrell et al., 2021), providing a good representation of the most abundant
114 diatom species in the ocean (Malviya et al., 2016). Finally, 54 single-cell and metagenome-
115 assembled genomes (sMAGs) from diatoms have been assembled from meta-transcriptome/-
116 genome data derived from *Tara* Oceans (Delmont et al., 2022), allowing complementary
117 insights into the biology and ecology of uncultured and uncultivable species (Nef et al.,
118 2022). Phylogenetic analyses of diatom genomes have revealed an extensive gene repertoire,
119 which can be considered in a phylogenetic sense to constitute a patchwork coming from an
120 ancient host, several endosymbionts acquired at different times, and bacterial horizontal gene
121 transfers (Dorrell et al., 2021, Vancaester et al., 2020). Reflecting their complex evolutionary
122 histories and phylogenetic distance (perhaps a billion years) from better-studied model
123 eukaryotes within the animals, fungi and plants, the functional organization and evolutionary
124 trajectories of diatom genomes are highly distinctive. These include families of novel
125 transposable elements (Hermann et al., 2014), novel epigenomic marking of chromatin
126 (Veluchamy et al., 2015; Zhao et al., 2021) and an apparent lack of structured centromeres
127 (Bowler et al., 2008).
128
129 Combined functional genomic approaches in model species have been used to begin to
130 decipher molecular actors regulating diatom physiology and distinct cellular and metabolic
131 features. These include extensive energetic exchanges between plastids and mitochondria that
132 augment CO_2 assimilation (Bailleul et al., 2015), a peculiar organisation of the photosystems
133 in the plastid membranes (Flori et al., 2017), and novel Light Harvesting Complex (LHC)
134 protein families (Bailleul et al., 2010; Buck et al., 2019). The central role of diatoms in
135 marine biogeochemical cycles has further been explored through identification of molecular
136 mechanisms of nutrient uptake and metabolism, e.g., for silica (Nemoto et al., 2020), carbon
137 (Shen et al., 2017), iron (Gao et al., 2021), nitrogen (Rogato et al., 2015) and phosphorus
138 (Dell'Aquila and Maier, 2020). Comparative genomics and molecular physiology studies
139 have greatly contributed to predict the role of nearly half of the diatom gene repertoire

(Blaby-Haas and Merchant, 2019): as of July 2024, 6,781 genes from the nuclear genome (out of 12,357) have at least one annotation from Gene Ontology (GO), InterPro or Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. Notwithstanding, the functional significance of many other diatom genes remains largely unexplored.

Among all studied diatom species, the genomic resources for *P. tricornutum* are the most advanced (Russo et al., 2023). The first assembly of the *P. tricornutum* nuclear genome constituted 27.4 Mb and 10,402 total gene models, whose annotation was assisted by the availability of an extensive collection of Expressed Sequence Tags (ESTs) (https://mycocosm.jgi.doe.gov/Phatr2/Phatr2.home.html) (Bowler et al., 2008). At the same time, a complete mitochondrial genome (77 kb, containing 60 genes (Oudot-Le Secq and Green, 2011)) and plastid genome (117 kb, containing 162 genes (Oudot-Le Secq et al., 2007)) were assembled. Following the initial assembly, the nuclear genome annotation was refined using 90 RNA-Seq datasets and more advanced annotation algorithms, resulting in the Phatr3 annotation. This annotation is available on the Ensembl archive (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index; Rastogi et al., 2018). More recent annotations (e.g., using proteomic data (Yang et al., 2018) and a new telomere-to-telomere *P. tricornutum* genome assembly (using long read sequencing technology (Filloramo et al., 2021 and Giguere et al., 2022)) have been performed, but to date no new annotation has been proposed to the community. Concerning transcriptomics, 123 microarrays have been used to generate a co-expression network (Ashworth et al., 2016), which is deposited on the DiatomPortal website (http://networks.systemsbiology.net/diatom-portal), while gene clustering of RNA-Seq experiments were recently used to create the PhaeoNet database (Ait-Mohamed et al., 2020). The epigenomic PhaeoEpiView browser including published epigenomic data has also been generated (Wu et al., 2023). Comparative genomics including *P. tricornutum* with a complete set of functional annotations are also provided in the PLAZA diatom comparative genomics platform (Vandepoele et al., 2013, Osuna-Cruz et al., 2020).

Despite the wealth of information generated from *P. tricornutum* summarized above, the resources are often disconnected, not always updated, or are incomplete, which diminishes their potential to yield valuable insights to the research community. To improve the connectivity between all the published omics data, we present here DiatOmicBase, a gene-centered web portal aiming to centralize all omics data related to diatoms. By focusing on the *P. tricornutum* model system, we provide different examples of the utility of this resource,

e.g., from the analysis of gene and protein characteristics, as well as gene expression analyses under different conditions using previously published datasets. We demonstrate how querying the DiatOmicBase resources can enable a deeper understanding of diatom gene products and their roles in specific physiological and cellular processes, and whole cell metabolic fluxes that may be exploited for biotechnological and synthetic biology applications (Kumar et al., 2022). In addition to exploring existing datasets, the portal also allows submission of a user's own data to provide a platform for common analyses to be performed.

Results and Discussion

Gathering information to develop a complete database centered on P. tricornutum genes

The DiatOmicBase website offers a gene-centered approach to facilitate exploration of the functional roles and regulation of diatom genes using omics-based resources. Considering *P. tricornutum* as a primary example, 12,392 gene pages corresponding to the latest annotation of protein coding genes (12,357 from the nuclear genome, 35 from the chloroplast, 60 genes from the mitochondria) contain sequence information gathered on a genome browser (Figure 1), alongside general descriptions related to their functional annotation and evolutionary history (Figure 2).

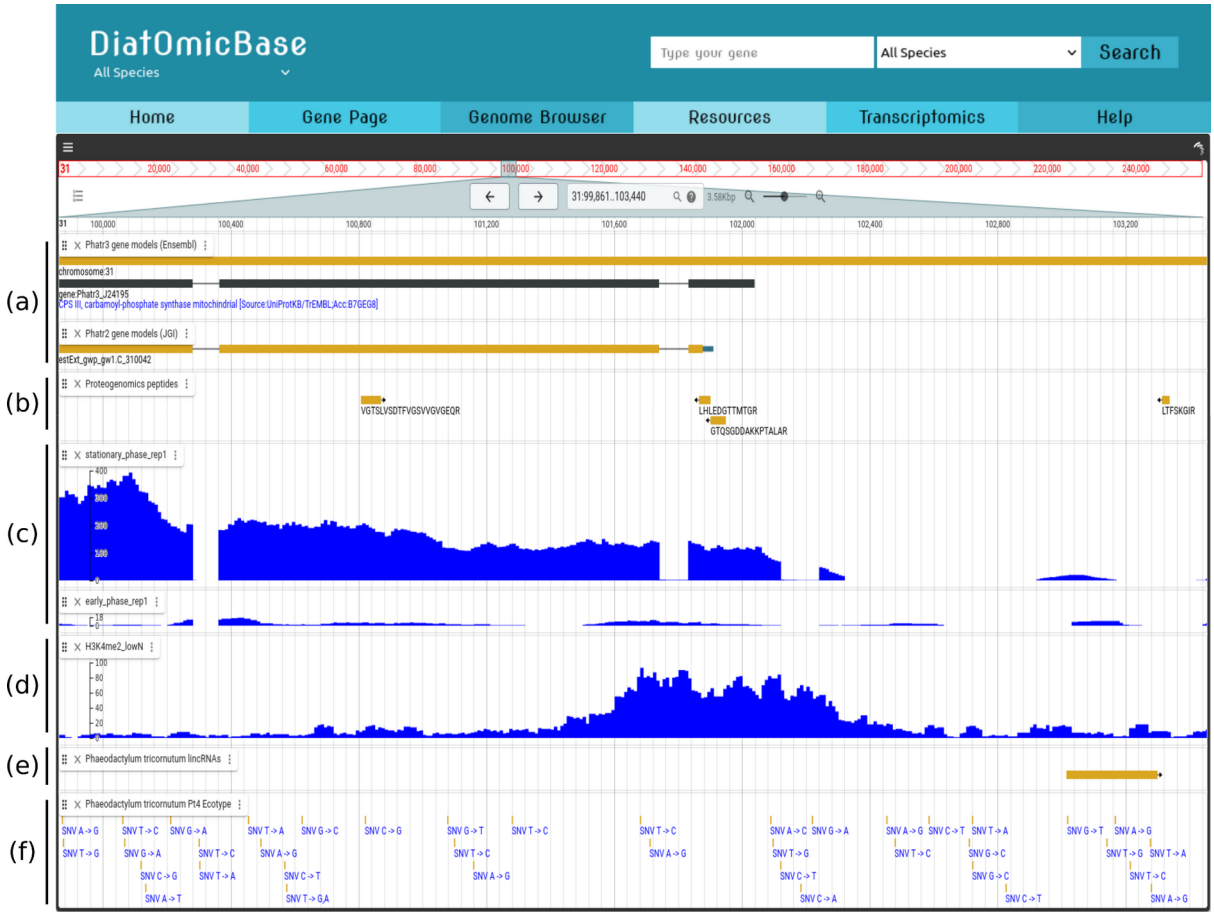


Figure 1. *Snapshot of the DiatOmicBase genome browser illustrating the different features used for an integrative analysis of the CPS III gene (Phatr3_J24195).*

(a) Phatr3 and Phatr2 gene models can be compared at different zoom levels.

(b) Zooming in on the 5' region reveals two peptide sequences mapped using a proteogenomic pipeline on an exon predicted by the Phatr3 gene model but not by Phatr2.

(c) Visualization of read mapping densities (here from Kwon et al., 2021) also validates the Phatr3 prediction.

(d) Density plots of histone marks (H3K4 methylation) under nitrate-depleted conditions.

(e) Long non-coding RNAs (Cruz de Carvalho et al., 2016)

(f) Single nucleotide variants from 10 different ecotypes (Rastogi et al., 2020). This figure specifically shows only the Pt4 variant.

DiatOmicBase uses the 27.4 Mb **assembly** obtained in 2008 (Bowler et al., 2008) from the sequencing of *P. tricornutum* accession Pt1 8.6 (deposited as CCMP2561). As the latest published **gene prediction**, **Phatr3** (Rastogi et al., 2018) is defined as the standard in the website. This reannotation was obtained using existing gene models, expression data and

protein sequences from related species to train prediction programs and predicts 12,089 gene models. To preserve the continuity of research by conserving gene identifier correspondence, the previous gene annotation, **Phatr2** (Bowler et al., 2008), comprising 10,402 gene models, is also available considering that several genes have been manually annotated on this version by the diatom research community. Only 4,667 Phatr3 gene models display a perfect correspondence with Phatr2. The other Phatr3 gene models can be new (1,489), have a modified 5′ and/or 3′ (4,709), can be merged (194), split (262), antisense (346), or required manual curation (566).

The genome browser also displays the annotation based on **proteomics** data. Yang et al., (2018) analyzed the proteome of 45 samples from *P. tricornutum* grown under eight different conditions. The peptide sequences resulting from the protein digestion and mass spectrometry analysis are shown on the genome browser. Using a proteomic pipeline integrating a dedicated protein search database, Yang et al., (2018) confirmed 8,300 Phatr2 genes, and identified 606 novel proteins, 506 revised genes, and 94 splice variants that can all be found on the genome browser. Discrepancies in gene model prediction between Phatr2 and Phatr3 can be resolved either using these proteomic data, or by analyzing mRNA expression and possible splicing variants using the multiple transcriptomics datasets centralized in the DiatOmicBase genome browser (see case study below, Figure 1 and Figure S1). To allow visualization of within-species gene diversity beyond the Pt1 8.6 strain, the database also includes genomic data from ten *P. tricornutum* ecotypes, covering broad geospatial and temporal scales (Martino et al., 2007), that have been re-sequenced (Rastogi et al., 2020). Variant calling results including SNPs, small insertions and deletions, can be visualized on the genome browser.

In addition to the sequence-related features displayed on the genome browser (Figure 1), several other descriptions are available on each gene page (Figure 2 and Figure S2). Gene pages are accessible by querying the database using Phatr2 or Phatr3 identifiers, terms or identifiers for GO, KEGG and Interpro databases. Keyword searches are available on the general search bar and on the gene page search. Finally, gene pages can be retrieved using nucleotide, protein and translated nucleotide (blastx) BLAST, with the possibility to tune various parameters.

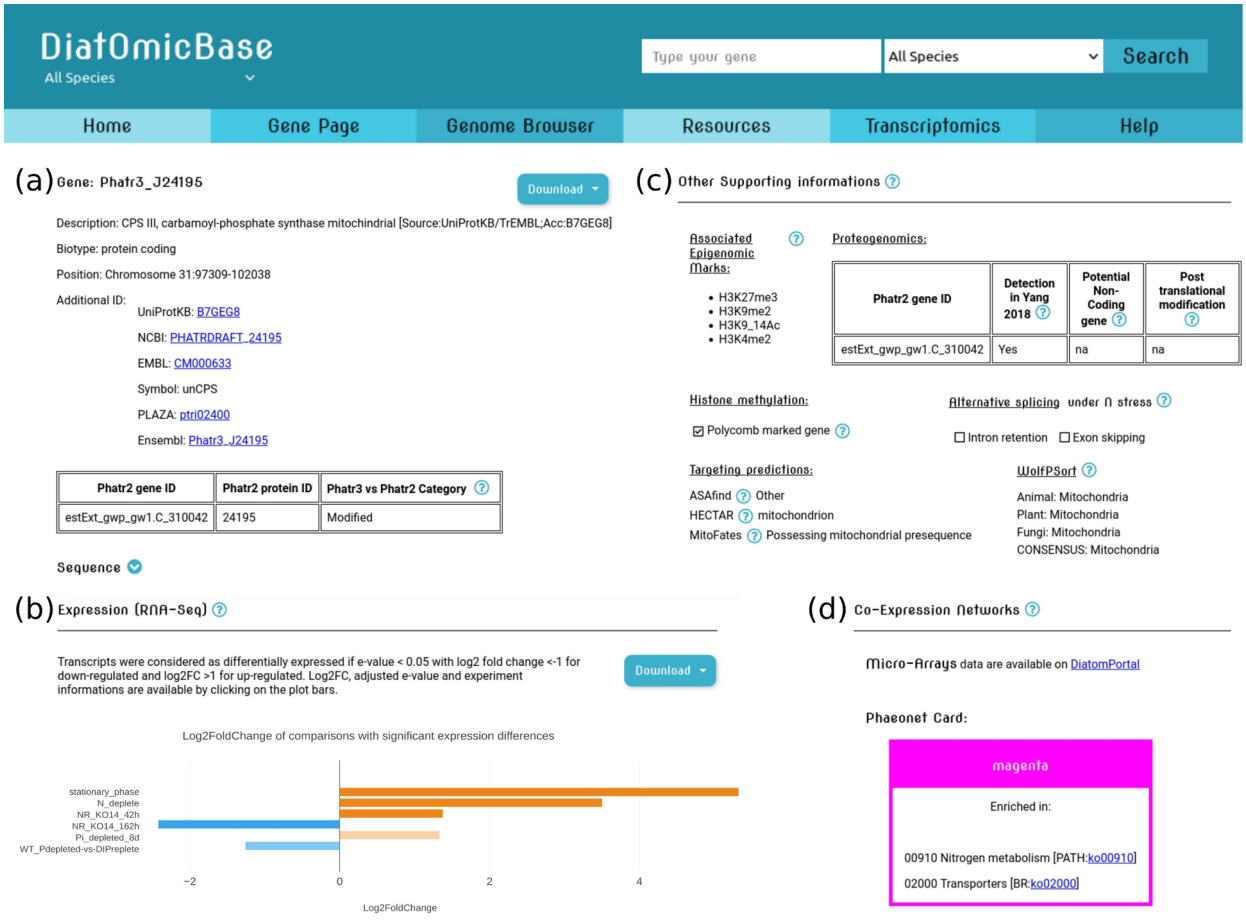


Figure 2. *Snapshot of DiatOmicBase gene page for CPS III.*

(a) The different gene identifiers and links are provided to facilitate navigation between databases.

(b) A barplot shows the log2 fold change of several RNA-Seq comparisons. The "stationary_phase" comparison contrasts early and stationary growth phases (Kwon et al., 2021). "N_deplete" compares nitrogen-free and nitrogen-replete samples after 48 [1] hours of treatment (Levitan et al., 2015). NR_KO14 refers to a transgenic knockout line for the nitrate reductase NR gene compared to wild-type cultures. In this experiment cells were grown in N free media and subsequently resuspended in NO3- for 42 and 162 hours (McCarthy et al., 2017). "Pi_depleted_8d" compares phosphate (Pi) replete cultures with cultures after 8 days of Pi depletion (Cruz de Carvalho et al., 2016). "WT_Pdepleted-vs-DIPreplete" compares wild-type strains grown in Pi depleted medium and replete with Dissolved Inorganic Phosphorus (DIP (Li et al., 2022).

(c) The "Other supporting information" paragraph summarizes information collected from different studies (explanations and references are provided using the

interactive "question mark"). Notably, HECTAR and Mitofates predict the expression of CPS III in the mitochondrion.

(d) To further integrate functional data, a link is provided to access the corresponding microarray-based co-expression network and the Phaeonet clusters (based on RNA-Seq data).

A snapshot of the complete page is available as Figure S2.

The **gene annotations** include identifier correspondences with UniprotKB and NCBI, retrieved from their respective databases, and the former gene annotation Phatr2 provided in Rastogi et al., (2018). Annotations from the GO project (Ashburner et al., 2000) were retrieved from UniprotKB (The UniProt Consortium, release 2021_01). GO terms for functional analyses were generated via automatic annotation, and cover three domains: the Subcellular Component where the gene products are localized, the Molecular Function informing the main activities of the gene product, and the Biological Process, the set of molecular events involving the gene product. Several domain and protein family predictions can be retrieved from UniprotKB, integrating 20 different databases (e.g., InterPro, Pfam, Gene3D, TIGRFAMs, CDD; the complete list is available here). These databases integrate different automated and/or manually curated protein signatures to ease the identification of protein functions. Functional orthologs, as previously reported in Aït-Mohamed et al., (2020) using the KEGG orthology database (Kanehisa, 2002), provide insights about molecular functions using hierarchically-structured biochemical pathways.

Transposable Elements (TEs) represent around 75% of the detected repetitive elements in the Phatr3 genome annotation (Rastogi et al., 2018; Giguere et al., 2021). With their ability to insert into genes or regulatory sequences, TEs act as key players in the organisation and expression of the genome, contributing to phenotypic diversity and, ultimately to the species evolution (Abbriano et al., 2023). A specific track has been created on the DiatOmicBase genome browser to visualize their positions by transforming the flat tables produced in Rastogi et al., (2018) into gff files. The majority (2,790, ~75%) of the TEs are associated with epigenetic marks (see below), which can also be important regulators of gene expression (Figure S2).

Epigenetic modifications play a pivotal role in regulating gene expression, influencing development, adaptation to environmental changes, and maintaining genome stability. In *P.*

tricornutum, whole genome methylation (Veluchamy et al., 2013) and the distribution of five histone marks (Veluchamy et al., 2015) have been described for the most used *P. tricornutum* strain, Pt1 8.6 (CCMP2561) grown in standard culture conditions. The methylome was obtained by digestion of three replicate DNA samples with the methyl-sensitive endonuclease McrBC followed by hybridization to a 2.1-million-probe McrBc-chip tiling array of the *P. tricornutum* genome (Veluchamy et al., 2013). The 3,950 methylated regions shown on the genome browser result from normalization on these three biological replicates. Five histone marks were chosen as they are known to be involved in transcriptional activation or repression: H3K4me2, H3K9me2, H3K9me3, H3K27me3 and H3AcK9/14. Two biological replicates of each histone modification were analyzed using ChIP-Seq, resulting in the discovery of 119,000 regions annotated with a set of chromatin states covering almost 40% of the genome. Following this first description of a diatom epigenomic landscape, changes have been examined in response to nitrate depletion, both by analyzing histone modifications (H3K4me2, H3K9/14Ac and H3K9me3 using Chip-seq) and DNA methylation (bisulfite deep sequencing) (Veluchamy et al., 2015). These marks are also displayed on the genome browser. DiatOmicBase also provides a direct link to the PhaeoEpiView epigenome browser which include epigenomic data (mostly DNA methylation and post-translational modifications of histones) on the newly assembled genome (Filloramo et al., 2021).

Different kinds of **small noncoding (sn)RNAs** (25 to 30 nt-long) have been described in P. tricornutum (Rogato et al., 2014) after sequencing of short RNA fragments isolated from cells grown under different conditions of light and nutrients. Their sequences have been mapped onto the genome browser. The majority of snRNAs map to repetitive and silenced TEs marked by DNA methylation and recent evidence indicates their role in the regulation of epigenetic processes (Grypioti et al., 2024). Other snRNAs target DNA-methylated protein-coding genes, or are derived from longer noncoding RNAs (tRNAs and U2 snRNA) or are of unknown origin. Long noncoding (lnc)RNA sequences have been shown to play a significant role in transcriptional mechanisms and post-translational modifications (Statello et al., 2021; Mattick, 2023). Although most of the well characterized lncRNAs stem from mammalian systems, recent work has shown that marine protists, including diatoms, all express lncRNAs (Debit et al, 2023). Among the different categories of lncRNAs, over 1,500 lincRNAs (intergenic lncRNAs) and ~3,200 lncNATs (antisense lncRNAs), that have previously been predicted from transcriptome mapping to the P. tricornutum genome (Cruz de Carvalho et al.,

2016; Cruz de Carvalho & Bowler, 2020), are likewise available on the genome browser, but are not integrated on the gene pages.

Links to a range of already existing web databases for **comparative genomics** are also provided. Information can be mined using PLAZA Diatoms (Osuna-Cruz et al., 2020) that includes structural and functional annotation of genome sequences derived from 26 different species, including 10 diatoms. Complementary annotations, homologous and orthologous gene families, synteny information, as well as a toolbox enabling a graphical exploration of orthologs and phylogenetic relationships are available on the corresponding gene pages of PLAZA. Alongside this, **evolutionary history annotations** based on a ranked BLAST top hit approach obtained from 75 combined libraries from different taxonomic groups across the prokaryotic and eukaryotic tree of life have been generated (Rastogi et al., 2018).

Co-expression networks gather genes with similar expression patterns across samples, suggesting that they could be related functionally, regulated in the same way, or belong to the same protein complex or pathway. In *P. tricornutum*, two different studies of co-expression networks have been published. The first one, published by Ashworth et al., (2016) explored the hierarchical clustering of 123 **microarray** datasets generated from studies of silica limitation, acclimation to high light, exposure to cadmium, acclimation to light and dark cycles, exposure to a panel of pollutants, darkness and re-illumination, and exposure to red, blue and green light. A link to the corresponding gene page on DiatomPortal, the web platform containing this data, is available. More recently, Ait-Mohamed et al., (2020) performed Weighted Gene Correlation Network Analysis (WGCNA) of 187 publicly available and normalised **RNA-Seq** datasets generated under varying nitrogen, iron and phosphate growth conditions (Cruz de Carvalho et al., 2016; McCarthy et al., 2017) to identify 28 merged modules of co-expressed genes. The gene cluster identifier (denoted **PhaeoNet card**) and its KEGG pathway enrichments are indicated for each gene within these modules in DiatOmicBase. PhaeoNet modules for each gene are detailed in the supplementary information file available on the resource page.

Finally, supplementary information regarding the functions of each protein are provided: **post-translational modifications** extracted from the work of Yang et al., (2018); alternative splicing from the work of Rastogi et al., (2018) and also by analyzing the RNA-Seq data centralized in DiatOmicBase; *in silico* **targeting predictions** regrouping multiple different

363 predictive tools: HECTAR under default conditions (Gschloessl et al., 2008); ASAFind using
364 Signal P v 3.0 (Gruber et al., 2015; Dyrlov Bendtsen et al., 2004); MitoFates with cutoff
365 threshold 0.35 (Fukasawa et al., 2015); and WolfPSort with animal, plant and fungal
366 reference models (Horton et al., 2007), as previously assembled in Aït Mohamed et al.,
367 (2020).
368
369 Users have the possibility to leave comments on each gene page, indicating a reference and
370 one or more predefined labels to facilitate indexing (Figure S2). This tool is expected to help
371 the community to centralize information that is not easily accessible or cannot be
372 automatically retrieved, such as the availability of transgenic lines, or evidence for
373 transcriptional regulation or alternative splicing. A peer-reviewed reference is mandatory and
374 comments can be moderated by the DiatOmicBase committee. The comments can also be
375 used to specify a correct gene model when different gene annotations (Phatr2 and Phatr3) are
376 not consistent.
377
378 ***Transcriptomic Data and Analysis***
379 We collected raw data available at NCBI from the RNA-Seq short reads BioProjects
380 published from *P. tricornutum* (different ecotypes and selected transgenic lines), exposed to
381 different conditions. These kinds of studies cover nutrient limitation (nitrate, phosphate, iron,
382 vitamin B12), responses to chemical exposure (decadenial, naphthenic acids, glufosinate-
383 ammonium, L-methionine sulfoximine, rapamycin, and nocodazole), different CO_2 and light
384 levels, response to grazing stress or competition. Samples involving transgenic lines (nitrate
385 reductase and aureochrome photoreceptor knock-outs, alternative oxidase and cryptochrome
386 knock-downs, chitin synthase transgenic cell lines, *etc*.) were compared to the corresponding
387 wild-type samples. Morphotype-related transcriptomes were all compared together as well as
388 growth stage transcriptomes. When studies consisted of time series, sample sets were
389 compared in control *versus* treatment pairs for each timepoint but not between timepoints. As
390 of July 2024, DiatOmicBase includes 33 studies, encompassing 1,398 samples and 135
391 comparisons.
392
393 For each BioProject, the most relevant pairwise comparisons were chosen to assess gene
394 expression regulation in the different sample sets using the R package DESeq2 (Love et al.,
395 2014). Bar plots showing up- and down-regulation illustrate the results of the pre-computed

comparisons on each gene page. Only significant comparisons are graphically shown but the list of non-significant comparisons and samples without read matches are provided.

On the transcriptomics page, users can also reanalyze public Bioprojects, defining the samples to compare (see case study 2). For public data, the gene-level read counts are already computed for each sample, and users only have to select the samples to compare. Moreover, users can also analyze their own data (see case study 3). For private data, inputs are a gene-level read count table or any equivalent expression matrix and a table informing how the samples should be grouped to be compared.

Gene expression analyses can subsequently be performed using the web application "integrated Differential Expression and Pathway analysis" (iDEP; Ge et al., 2018). Connecting several widely used R/Bioconductor packages and gene annotation databases, iDEP provides a user-friendly platform for comprehensive transcriptomics analysis, including quality control plots, normalization, PCA, differential expression analysis, heatmaps, pathway and GO analysis, KEGG pathway diagrams, functional annotation, co-expression networks, interactive visualizations, and downloadable gene lists. iDEP enables pairwise comparisons or more complex statistical models including up to 6 factors. Co-expression networks can be examined using WGCNA. All the analysis steps are customisable; methods and parameters can be easily tuned using dialog boxes.

Finally, expression patterns of several genes can be analyzed by drawing customized heatmaps (Figure 3). In this case the user can provide a gene list and select the sample comparisons to be shown on the plot.

Case studies

Improving gene model prediction and visualization of gene expression and regulation

DiatOmicBase can aid functional analyses of diatom genes by improving prediction of protein-coding genes and their regulation. In Figures 1 and 2, we show the information that can be readily obtained in DiatOmicBase, using as an example the *P. tricornutum CPS III* gene, encoding a carbamoyl phosphate synthase involved in the urea cycle. In diatoms, the discovery of an ornithine-urea cycle (OUC) was unexpected as it was previously thought to be specific to animals, to allow removal of excess $NH4^+$ derived from a protein-rich diet.

Rather than eliminate NH_4, diatoms have been proposed to conserve this precious resource by using the pathway to cope with fluctuations in nitrogen availability in the ocean (Allen et al., 2011, Smith et al., 2019), maintaining the cellular balance of carbon and nitrogen. The key enzyme of the OUC, carbamoyl phosphate synthase (CPS) is encoded in the *P. tricornutum* genome by two copies. One copy, named *unCPS* or *CPS III* (Phatr3_J24195) and using ammonium as a substrate, is predicted both by HECTAR and MitoFates to be targeted to the mitochondria (Figure 2c), in agreement with the mitochondrial localization observed by microscopy (Allen et al., 2011). Furthermore, a *T. pseudonana* homologue of *unCPS* (Thaps3a_40323) has been recovered from proteomic data derived from purified mitochondrial fractions. A second *P. tricornutum* paralog, pgCPS2 (encoded by Phatr3_EG01947) has been inferred to be cytosolic.

In DiatOmicBase, the peptide alignment from proteogenomic studies confirmed the Phatr3 gene model of *CPS III*, compared to the Phatr2 gene model (estExt_gwp_gw1.C_310042; Figures 1, 2 and S1), predicting a protein extended by 133 amino acids at the N-terminus, due to the translation from an earlier ATG initiation site in the genome (chromosome 31 position 102308 reverse strand, c.f. chromosome 31 position 101909 reverse strand). This gene model and also intron position were confirmed by analyzing RNA-Seq data from conditions where the *CPS III* gene is strongly expressed (Figure S1).

Considering quantitative gene expression trends, reanalysis of transcriptomic data in DiatOmicBase has shown that *CPS III* is highly expressed in cells experiencing three days of nitrogen deprivation (Levitan et al., 2015). Interestingly, its expression is also induced under phosphorus (P) depletion conditions, whereas Pi repletion of starved cells reverse this trend. These responses to Pi availability could be the result of the rapid cross-talk between Pi and N metabolism observed in diatoms (Helliwell et al., 2021). *CPS III* was also overexpressed in a nitrate reductase (NR) knockout line compared with the wild-type (wt) under conditions of nitrogen repletion (e.g., comparing the response of NR knockout versus wt in cells repleted with nitrogen for 42 h), but was down-regulated when cells became limited for this nutrient (e.g., NR KO versus wt after 162 h of nitrogen resupply) (McCarthy et al., 2017; Figure 2b). *CPS III* is also down-regulated in cell lines incubated in the complete absence of nitrogen compared with nitrogen-replete media for 4 and 20 hours (Matthijs et al., 2016), but was up-regulated when cells experienced prolonged nitrogen starvation (Levitan et al., 2015). It is therefore possible that nitrate uptake and internal cellular nitrate concentrations play a role in the hierarchical regulation of *CPS III* expression. We also note a dramatic increase in *CPS III*

expression in stationary-phase cells compared with exponential-phase cells, which may relate to functions in amino acid scavenging and recycling, but also in nutrient starvation responses. Information on *CPS III* gene expression in additional growth conditions and physiological states can be obtained by analyzing other centralized omics data (e.g., Figure S2).

The histone marks H3K4me2 and H3K9/14Ac, consistent with transcriptional activity, mapped to the 5' region of the gene both in nitrate replete and deplete conditions, suggesting that CPS III is also important in nitrate replete cells (Figure 2c). We furthermore note that *CPS III* groups in the PhaeoNet magenta card, a module of co-expressed genes that appear to be enriched in functions implicated in organelle amino acid and nitrate metabolism (Figure 2d). Other members of this module include the plastidial glutamate synthetase (Phatr3_J50912) and N-acetyl-gamma-glutamyl-phosphate reductase implicated in the plastidial ornithine cycle (Phatr3_J36913), as well as several other genes encoding proteins involved in nitrate assimilation. This might be consistent with a role of the urea cycle under nitrate replete conditions in recycling excess assimilated plastidial amines. Finally, while no TEs were found in the coding region, a lincRNA was detected in the 5' region of the gene. A dozen small RNAs were mapped onto the gene, and ecotypes displayed between 7 and 31 single nucleotide variants (Figure S2), suggesting further possible hierarchical factors influencing the function and evolution of this gene.

Identification of common or distinct regulators of diatom responses to light and nutrient variations

Here we show how transcriptomic resources centralized in DiatOmicBase can be exploited to perform novel comparative functional analyses across multiple datasets (i.e., meta-analyses) and derive new information about common or specific regulators possibly implicated in diatom acclimation to environmental cues. Fluctuations in nutrient availability are recurrent in the marine environment, with nitrogen (N) and phosphorus (P) being amongst the main limiting nutrients for primary productivity. Both nitrogen and phosphorus deficiencies lead to a reduction in photosynthetic activity and a halt in cell division in diatoms (Jaubert et al., 2022; Levitan et al., 2015; Cruz de Carvalho et al., 2016; Matthjis et al., 2016). This may result in a cellular quiescent state, which is reversed when nutrients are resupplied (Cruz de Carvalho et al., 2016; Matthjis et al., 2016; Dell'Aquila and Maier, 2020). On the other hand, light is the major source of energy for photosynthesis, and a critical source of information

from the external environment. In recent years, the exploration of diatom genomes and of transcriptomic data obtained from cells exposed to different light conditions (wavelength, intensity and photoperiod cycles) and targeted functional analyses of selected genes in *P. tricornutum* has identified peculiar regulators of diatom photosynthesis (Lepetit et al., 2022), novel photoreceptors responsible for light perception, and an endogenous regulator controlling responses to periodic light-dark cycles (Jaubert et al., 2022). However, the regulatory cross-talk between light and nutrient signaling pathways remains poorly understood in diatoms. To this aim, here we reanalyzed already available RNA-Seq data from cells experiencing different light conditions (high light stress, fluctuating light, different colours) as well as different nitrate and phosphate depletion and repletion conditions.

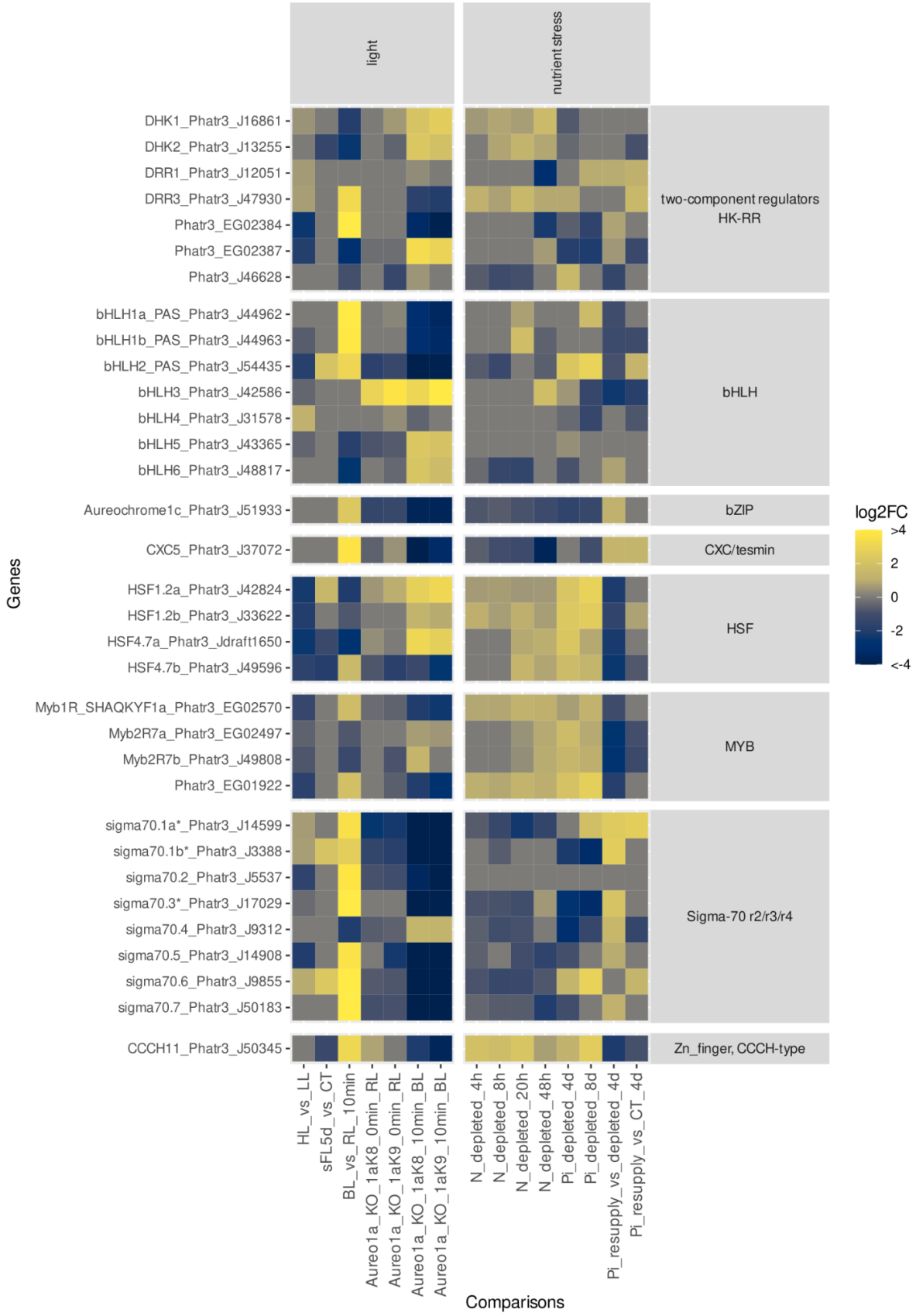
In Figures 3, S3 and S4, we focused on gene expression changes for two-component regulators and transcription factors (TFs) encoded in the *P. tricornutum* genome. These regulators orchestrate the physiological response(s) to a given stress by participating in signaling cascades through differential expression regulation. The genome of *P. tricornutum* contains a high number of bacterial-like two-component sensory histidine kinases (Bowler et al., 2008), defined on the basis of the presence of a histidine kinase domain (PF00512) and/or a response regulator domain (PF00072). In addition, the histidine phosphotransmitter protein (Hpt), which was reported missing in the first version of the genome (Bowler et al., 2008), was later identified in Phatr3 (Phatr3_J33969, PF01627) and was included in the analysis (Figure 3). Regarding two-component regulators, we observed significant gene expression changes for some of these regulators following a red to blue light shift (e.g., *DHK1*, *DHK2*, *DRR3*, *EGO2384*, *EGO2387*). The loss of these responses in two independent Aureochrome blue light receptor KO lines (Mann et al., 2020) compared to wild-type cells indicates that this blue light photoreceptor participates in the regulation of their expression. By contrast, *Phatr3_J46628* is not affected by any changes in light conditions, but is over-expressed under prolonged phosphate depletion (4d) (Figure 3). Interestingly, *DDR3* is found to be up-regulated under blue-light treatment, but also under both N (from 4h to 48h) and Pi depletion (4 days) as well as under Pi resupply (Figure 3 and S3), but not under HL. It therefore appears likely that a blue light-activated signaling cascade participates via DDR3 in the regulation of cellular responses to nutrient availability, but not in photoprotection. *Phatr3_EG02384* is expressed in the opposite fashion in response to HL and nitrogen depletion, suggesting a possible antagonistic role in light and nutrient signaling pathways. On

the other hand, *DRR1* appears not to be affected by light signals, but is specifically expressed under prolonged N deficiency and Pi depletion.

The analysis of TF expression patterns across treatments also provides interesting information (Figures 3 and S4). Between the genes encoding bHLH domain-containing proteins, the bHLH1a-PAS also known as RITMO1 shows a rapid induction by blue light treatment, which is repressed in *Aureo1* photoreceptor mutants, as previously shown (Mann et al., 2020). Neither HL treatments, nor short N and Pi deficiency affect its expression, as expected considering the role of this protein as an endogenous and robust timekeeper (Annunziata et al., 2019). Its expression and possible activity is, however, affected by prolonged nutrient stress treatments. bHLH2-PAS is differentially expressed under the tested light conditions as well as Pi deficiency, while bHLH3 appears to be expressed specifically under N and Pi deficiency. Interestingly, its expression is not induced by blue light, but is significantly affected in the *Aureo1a* mutants under red light. It is thus possible that *Aureo1a*, which is also a TF with a bZIP domain, can participate in regulation to nutrient changes, independently of its activity as a blue light sensor. A light-independent activity of *Aureo1a* is also supported by a recent study reporting altered rhythmic gene expression in *Aureo1* mutants compared to wild-type cells in constant darkness (Madhuri et al., 2024). Of the genes encoding sigma 70 factors, it is interesting to note that their expression is strongly modulated by HL or blue light, and for the three factors predicted to be plastid localized (sigma 70.1a, 1.b and 70.3*) also by Pi availability. This is consistent with a possible role for these factors in regulating plastid gene expression, which could also be sensitive to the physiological state of this organelle. Other TFs modulated by nutrient availability and blue light include Aureochrome1C of the bZIP family and Pt-CXC5 of the CXC/tesmin family, both of which are downregulated under nutrient stress and upregulated upon nutrient resupply, namely P, as well as under blue light (Figure 3). On the other hand, the opposite trend can be found for TFs of the HSF (HSF1.2a/b; HSF4.7a/b) and Myb families, which are also modulated by nutrient status and light, being up-regulated under nutrient and high light stress (Myb2R7a/b), as well as blue light (Myb1R_SHAQKYF1a; Phatr3_EG01922).

This analysis reveals the complex cross-talk between the regulatory pathways operating under light and nutrient fluctuations, which could enable opening new exploratory and functional research avenues aiming to shed light into the molecular processes underpinning diatom resilience to environmental stresses.

563



564

565

Figure 3. *RNA-Seq heatmaps of two components regulators and transcription factors*

Heatmap showing gene expression differences in light and nutrient stress experiments for different genes of interest. HL_vs_LL compares low light and high light acclimated wild-type (WT) strains (Agarwal et al., 2022). sFL5d_vs_CT compares strains acclimated to constant light conditions to severe light fluctuation conditions (Zhou et al., 2022). BL_vs_RL_10min compares WT strains under red light (RL) and 10 minutes after a shift to blue light (BL); In samples KO_1aK8_10min_BL and KO_1aK9_10min_BL, Aureo1a knockout K8 and K9 strains are compared to WT under RL, and after a 10 minutes shift to BL (Mann, et al., 2020).

N_deplete_ compares nitrate-deplete and control samples 4h, 8h and 20h after transfer to nitrogen-replete medium (Matthijs et al., 2016). N_deplete_48h compares nitrogen-free and nitrogen-replete samples after 48h of treatment (Levitan et al., 2015). Pi_depleted_4d and Pi_depleted_8d compare control samples and cultures without phosphate supplement during 4 and 8 days. Pi_resupply-vs-depleted_4d and Pi_resupply-vs-CT_4d compare samples with phosphate re-supplementation after 4 days of starvation (Cruz de Carvalho et al., 2016).

** based on the prediction supported by the database, these genes are localized in the plastid.*

DiatomicBase for the *de novo* analysis of previously unpublished RNA-Seq data

In DiatomicBase, we have also developed the possibility for users to analyze their own RNA-Seq data. Here, we show the analyses of new data obtained from *P. tricornutum* lines over a progressive two-week iron (Fe) starvation time-course. More specifically, we compared gene expression changes between cell lines grown in Fe replete media and those experiencing short-, medium- and long-term Fe withdrawal conditions (Figure S5).

Principal Component Analysis of the RNA-Seq data, performed with the integrated iDEP platform in DiatOmicBase, revealed three groups, corresponding to Fe-replete, short-term (3 days), and medium/ long-term (7, 14 days) Fe limitation (Figure 4a). Calculation of the k-means revealed four distinct clusters enriched in different GO terms that were differentially regulated in response to different periods of Fe limitation (Figure 4b). The first cluster (generated as Cluster A by the iDEP calculations) was strongly induced after 3 days Fe withdrawal relative to Fe-replete conditions and was significantly ($P< 10^{-15}$) enriched in genes encoding proteins with functions relating to photosynthesis (photosynthesis light reactions,

protein-chromophore linkage, generation of precursor metabolites and energy), alongside translation and peptide biosynthesis-associated processes (Figure 4c). This might relate to a transcriptional and translational upregulation of light-harvesting protein complexes in particular to compensate for diminished Fe availability limiting the synthesis of photosystem I (Gao et al., 2021). This contrasted with a second cluster (Cluster C) that was uniquely downregulated in short-term Fe limitation conditions, and showed a weaker ($P < 0.001$) enrichment in photosynthesis and transcriptional regulation processes but no other enriched GO terms (Figure 4b). The final two clusters, Clusters B and D, were strongly up- and downregulated, respectively, by medium and long-term Fe limitation compared to short-term and Fe-replete conditions. These showed weak enrichments ($P < 0.0001$) in transport processes, i.e., ion and organic metabolite transport (Figure 4b). These may relate to the induction of Fe uptake systems in response to prolonged Fe withdrawal, alongside changes in the internal metabolite profiles and organic acid transport activities of Fe-limited cell lines (Gao et al., 2021).

While the data from DiatOmicBase provide insights into transcriptomic changes, they can be used to inform user construction of more complex evaluations of the links between gene expression and function. As an example of this, in Figure 4c we present Volcano Plots for DEGs generated from the DiatomicBase site, alongside tabulated fold-change and P-values calculated for genes known to be involved in Fe stress metabolism (Figure 4d; Gao et al, 2021). Concerning known Fe-stress associated proteins, many were already strongly upregulated after three days Fe limitation compared to Fe-replete lines, confirming the immediate impacts of Fe withdrawal on cell physiology (Figure 4c) (Gao et al., 2021). The most significantly upregulated of these genes at all three time points (P-value < 10^{-100} for 3, 7 and 14-day Fe limitation) encoded a plastidial cytochrome c_6 (Phatr3_J44056), which mediates photosynthetic electron transfer between cytochrome b_6f and photosystem I, underlining the importance of photosystem remodelling in both short-term and sustained Fe limitation (Allen, de Paula et al., 2011). We note that *P. tricornutum* does not possess an endogenous plastocyanin gene (Groussman, Parker et al., 2015) that could be induced to compensate for cytochrome c_6 under Fe-limitation.

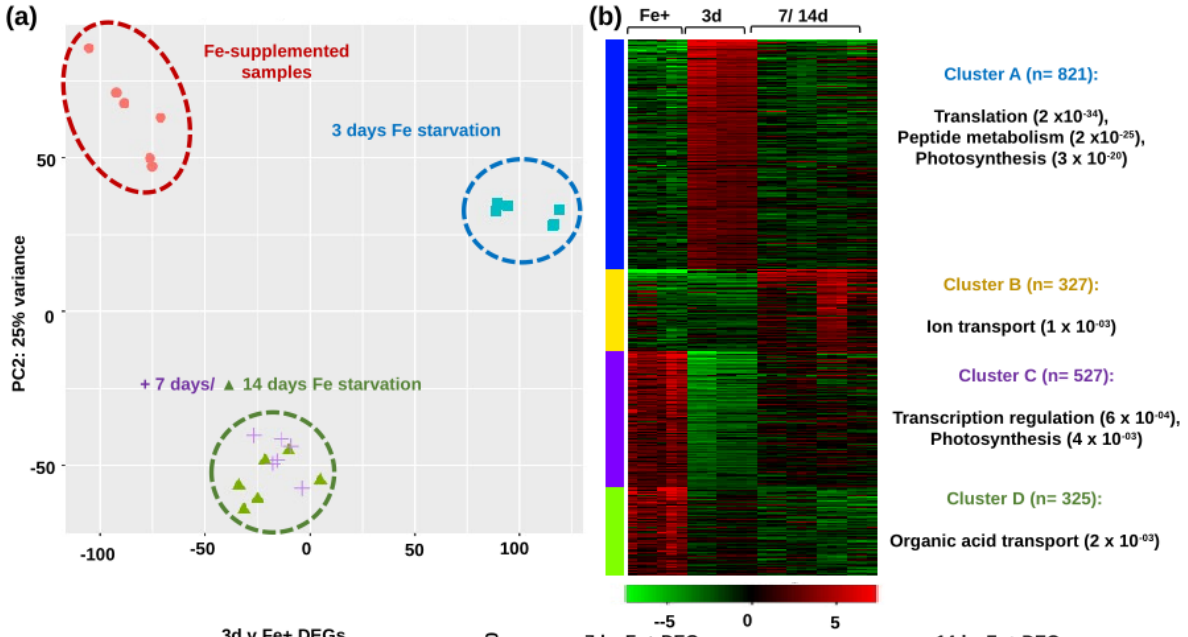
Consistent with previous studies, we see the upregulation of known iron-stress related genes at all Fe-limitation time points studied. Constitutively upregulated genes include *ISIP2a/ phytotransferrin* (*Phatr3_54465/ Phatr3_54987*) and *ferric reductase 2* (*Phatr3_J54982*),

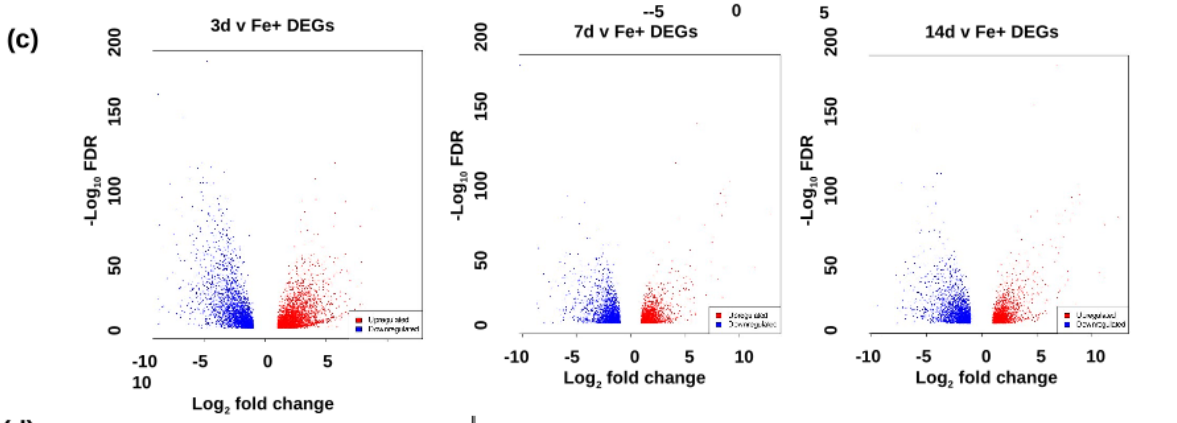
implicated in the reductive uptake of Fe from the cell surface (Morrissey, Sutak et al., 2015, McQuaid, Kustka et al., 2018); alongside genes encoding the newly identified plastid-targeted partners of ISIP2a *pTF.CREG1* (*Phatr3_J51183*), *pTF.ap1* (*Phatr3_J54986*), and the gene encoding the plasma membrane-located ISIP2a-binding protein pTF.CatCh1 (*Phatr3_J52498*) (Allen, Laroche et al., 2008, Turnšek, Brunson et al., 2021). These results suggest that reductive Fe uptake in diatom cells is relevant to both short- and long-term Fe limitation. We also observed upregulation at all time points for *ISIP3* (*Phatr3_J47674*), encoding a protein of unknown function but proposed to participate in the intracellular trafficking or storage of iron (Behnke and LaRoche 2020, Kazamia, Mach et al., 2022); and flavodoxin 2 (*Phatr3_J23658*), which can diminish cellular iron requirements of flavodoxin in photosystem I (Setif 2001, Lodeyro, Ceccoli et al., 2012).

Our data also provide insight into Fe-stress associated diatom genes that show more distinctive responses to Fe limitation. For example, ISIP1 (*Phatr3_J55031*), implicated in the non-reductive transport of Fe-siderophore complexes from the cell surface to the chloroplast, shows no evidence of induction in the three day treatment, but strong induction in medium- and long-term limitation datasets (Figure 4c) (Kazamia et al., 2018), suggesting that it is induced more slowly than reductive iron uptake strategies. This might reflect the greater energetic cost or gene coordination required by diatoms to produce siderophores in response to Fe limitation, whose synthesis pathway remains unknown. Most dramatically, the gene encoding the proposed plastid iron storage protein ferritin (*Phatr3_J16343*) showed no response to either medium or long-term Fe-limitation but was significantly downregulated ($P < 10^{-08}$) in response to short-term Fe withdrawal. The role of ferritin in diatoms has historically been unclear, with some studies suggesting that it facilitates long-term Fe storage and tolerance of chronic starvation (Marchetti et al., 2009); and others that it may be transiently upregulated in response to Fe enrichment, allowing competitive removal of iron from the environment (Lampe et al., 2018, Cohen et al., 2018). Our data are broadly more consistent with the latter role for the *P. tricornutum* ferritin, although we note it is phylogenetically distinct to other diatom ferritins and may confer different physiological roles (Gao et al 2021). A greater similarity of this *P. tricornutum ferritin* gene with sequences from *Nannochloropsis* and ciliates than from other diatoms is also observed by analyzing PLAZA Diatoms.

Finally, our data provide some insights into the broader physiological responses that mediate short- and long-term diatom responses to Fe deprivation. The importance of light-harvesting

667 complexes for short-term responses to Fe-limitation (Figure 4c) is underlined by the strong
668 induction of genes encoding two Lhc proteins (Lhcf1, *Phatr3_18049*; Lhcf5, *Phatr3_J30648*)
669 that directly interact with one another in the PSI LHC (Joshi-Deo et al., 2010), although are
670 also both found in PSII (Gundermann et al., 2013; Nagao et al., 2021), and are implicated in
671 low-light adaptation (Gundermann et al., 2013). In contrast, we observed strong
672 downregulation of PGR5 (*Phatr3_J44748*), debatably implicated in cyclic electron flow
673 around PSI in diatoms (Grouneva et al., 2011, Johnson et al., 2014). It has recently been
674 proposed that *Chlamydomonas reinhardtii* and plant PGR5 indirectly participate in the
675 delivery of Fe to PSI, which might explain its immediate sensitivity to Fe depletion in our
676 data (Leister et al., 2022). In contrast, under long-term Fe-limitation but not under short- or
677 medium-term conditions, we identified dramatic ($P < 10^{-50}$) downregulation of the gene
678 encoding Aspartokinase (*Phatr3_J52316*), involved in lysine biosynthesis, as well as the
679 gene encoding Protoporphyrinogen oxidase (*Phatr3_J31109*), involved in the tetrapyrrole
680 branch of chlorophyll/haem biosynthesis, which may point to an overall quiescence of core
681 organelle metabolic pathways in response to sustained Fe deprivation (Allen et al., 2008; Ait-
682 Mohamed, Novák Vanclová et al., 2020).





(d)

Gene name	3d versus Fe+		7d versus Fe+		14d versus Fe+	
	Fold-change	-LOG10 P	Fold-change	-LOG10 P	Fold-change	-LOG10 P
Phatr3_J44056, cytochrome c6	5,71	119,73	6,17	139,73	6,80	170,35
Phatr3_J54465, ISIP2A	5,02	33,11	8,04	83,82	6,32	51,74
Phatr3_J54987, ISIP2B	6,53	20,11	12,95	76,94	12,39	70,58
Phatr3_J54982, Ferric reductase 2	2,41	6,44	2,63	7,24	4,68	23,33
Phatr3_J51183, pTF.CREG1	4,29	23,73	8,62	93,70	8,14	83,69
Phatr3_J54986, pTF.ap1	4,03	19,75	8,65	89,26	8,74	91,14
Phatr3_J52498,pTF.CatCh1	3,39	12,64	8,46	75,98	8,03	68,53
Phatr3_J47674, ISIP3	4,76	30,36	8,31	91,00	7,53	74,78
Phatr3_J23658, Flavodoxin 2	4,49	24,28	9,16	99,18	8,21	79,80
Phatr3_J55031, ISIP1	N/A	N/A	11,48	39,42	10,64	34,05
Phatr3_J16343, Ferritin	-4,11	8,82	N/A	N/A	N/A	N/A
Phatr3_J18049, Lhcf1	5,70	83,76	3,15	25,14	1,76	8,06
Phatr3_J30648, Lhcf5	8,73	86,86	4,57	23,53	3,38	13,07
Phatr3_J44748, PGR5	-4,22	104,96	-2,24	30,27	-2,39	34,19
Phatr3_J52316, aspartate kinase	-1,90	17,63	-1,60	12,26	-4,18	79,05
Phatr3_J31109, protoporphyrinogen III oxidase	-4,24	44,71	-2,93	21,67	-5,81	66,89

Figure 4. de novo RNA-Seq analysis of iron starvation response.

Comparison of gene expression from cells adapted to 12h light:12h night cycles under 19°C, and either Fe-replete ("Fe+"), 3 day ("FeS"), 7 day ("FeM") or 14 day ("FeL") Fe-limitation. A schematic experimental design is provided in Figure S5.

(a) PCA of RNA-Seq Transcripts Per Million (TPM) values calculated with the integrated iDEP module in DiatOmicBase. Three distinct treatment groups are visible, suggesting relative equivalence of FeM and FeL treatments.

(b) k-means clusters, showing selected enriched GO terms (P < 0.001), calculated with the iDEP module in DiatOmicBase. Four clusters with distinctive biological identities show different relationships to both short-term and medium/ long-term Fe-limitation.

(c) Volcano plots of DEGs inferred with iDEP with threshold P-value 0.1 and fold-change 2.

(d) Tabulated fold-changes and -log10 P-values of genes associated with Fe stress metabolism, following Gao et al., (2021).

Perspectives

DiatOmicBase provides a comprehensive database and analytic modules integrating genomic, epigenomic, transcriptomic and proteomic data from published diatom genomes. In this work, we focused on the resources centered around *P. tricornutum*. The first objective of this centralized database is to provide the community with a powerful tool for assessing correct gene models. Automated gene annotation remains error-prone, especially considering the large proportion of diatom genes whose function is unknown and which have no clear homologues in other systems. The peptide and RNA/DNA sequence mapping provided in DiatOmicBase may help to identify the correct form, as shown for the *CPS III* gene; DiatOmicBase makes provisions for user annotation and correction of individual gene pages. Given that *P. tricornutum* is the diatom species with the largest collection of genomic resources and data, the correct gene annotation in this species also represents a powerful support for correct gene prediction in other diatom species and from environmental data.

Analyzing the expression of genes in different circumstances has been facilitated by grouping results of several RNA-Seq experiments involving different conditions. In the case studies described in this work, we have shown how our tools can help to identify common or specific regulators of responses to various light changes and nutrient stress conditions. This offers new opportunities to characterize the still largely unknown signalling pathways involved in the perception and acclimation to complex changing environments. Using the iDEP pipeline,

public RNA-Seq datasets can be re-analyzed, enabling to answer questions different from the ones in the original articles. In order to enrich our resources, we would like to encourage the community to inform us when new data becomes available.

DiatOmicBase also facilitates the study of gene functions, essential for deciphering the physiology and metabolic capacities of these microalgae and harnessing their potential for biotechnology and synthetic biology Assigning a phenotype to a protein requires overcoming many challenges (Heydarizadeh et al., 2014) as automated annotations should be validated by biochemical evidence and comparative analyses between wild-type and mutant lines. Metabolic engineering and synthetic biology resources in diatoms are rapidly evolving (Russo et al., 2023), yet they are limited by some aspects of diatom metabolism that are not yet well understood. For example, we still do not know how many metabolic pathways, including those that are important for biotechnology such as the biosynthesis of carotenoids and terpenoids, react to environmental conditions. Nor do we know the mechanisms by which these pathways are regulated, or the subcellular location of the different enzymes. Protein targeting predictions, expression data available or newly generated and analyzed in DiatOmicBase, in conjunction with metabolomics data can be exploited for strain engineering strategies, by improving gene-enzyme-function associations, identifying unknown enzymes, and reveal still elusive aspects of pathway regulation. For designing pathway engineering and optimizing cultivation strategies, DiatOmicBase can be particularly useful for investigating co-regulation of industrially-relevant metabolic pathways or key pathway nodes with transcription factors (Figure 3) across various conditions to identify regulators that can control entire metabolic pathways. These are primary genetic targets for strain engineering strategies which are promising and feasible (Song et al., 2023), but also remains a largely untapped strategy in diatom biotechnology.

The future development of the DiatOmicBase should include improved genome assemblies and annotations. Unfortunately, *P. tricornutum* genome complexity (especially TEs) has prevented significant improvement of the reference genome obtained by Sanger sequencing. A new assembly using the latest technologies of long read sequencing has recently been published (Filloramo et al., 2021). However, this long-read derived assembly still lacks continuity resulting in more than 200 scaffolds (while the Sanger assembly comprised 88 scaffolds with 33 chromosome-level scaffolds). As no gene annotation was associated with this assembly, it was not possible to fully exploit this new resource in this first version of

DiatOmicBase, but we aim to include it following it's annotation. As shown for the *CPS III* gene, DiatOmicBase should help to resolve conflicting gene model predictions derived by new sequencing and assembly projects. The database makes provisions for user annotation and correction of individual gene pages.

The addition of new diatom species in DiatOmicBase is also planned, especially those with the most developed omic resources. As of September 2024, preliminary DiatOmicBase portals are available for the model centric species *T. pseudonana* (https://www.diatomicsbase.bio.ens.psl.eu/genomeBrowser?species=Thalassiosira+pseudonana) and the model sexual diatom *P. multistriata* (https://www.diatomicsbase.bio.ens.psl.eu/genomeBrowser?species=Pseudo-nitzschia+multistriata). Users can automatically search for gene information in these different species. Further developments will include the addition of new proteomic data, the incorporation of comparative genomic and automated phylogenetic analyses of individual genes, functional genomic information from transgenic lines, and precomputed biogeographical distributions of environmental homologues of individual diatom genes from *Tara* Oceans (Vernette et al., 2022).

Experimental procedures

Website architecture

The back-end server of the website consists of an API server coded with the FastAPI framework. It includes a local PostgreSQL database that is accessed using the SQLAlchemy library. Data from NCBI/EBI/Ensembl are fetched and inserted or updated in the local database using a python loader script. The genome browser used is Jbrowse 2 (Buels et al., 2016). A R-shiny iDEP instance (Ge et al., 2018), hosted on the local server, was modified to contain *P. tricornutum* genome annotation and to automatically load public data from DiatOmicBase. A background worker coded in python is used to run more computationally intensive user calculations (Blast or differential expression analysis) and ensures that these do not use more resources than allowed, using a queuing system. The possibility to comment gene pages used Commento widgets. To ensure data privacy, Commento and its postgreSQL database are self-hosted.

The front-end part is developed with the React framework and uses static rendering with Next.js for performance. The FastAPI and Next.js instance communicates with the user through an Apache proxy to retrieve user requests and display results. Cite the Github address

Analysis of publicly available transcriptomic data

We collected raw data available at NCBI from the RNA-Seq short reads BioProjects published from *P. tricornutum* (different ecotypes and selected mutants), exposed to different conditions. Short read sequences were 35 to 150 bp long, principally generated from Illumina or DNBSeq. Fastq files were analyzed using the Bioinformatic pipeline nf-core/rnaseq (Patel et al., 2021). Briefly, raw reads were cleaned and merged before being mapped with Hisat2 on their reference genome and quantified using featureCounts (Liao et al., 2014), an algorithm specifically designed to quickly and efficiently quantify the expression of transcripts using RNA-Seq data. Read coverage files are displayed on the genome browser of each gene page.

For each BioProject, pairwise comparisons were chosen to assess gene expression regulation in the different sample sets using the R package DESeq2 (Love et al., 2014). All the replicates deemed to be available and of good quality (based on mean quality score computed with FastQC) were used to estimate biological and technical variation. In studies investigating transcriptomic responses to environmental variations, the tested conditions were compared with their respective control groups. In cases involving mutant lines, precomputed comparisons were made between mutants and wild-types.

Gene expression analyses can be performed with the web application “integrated Differential Expression and Pathway analysis” (iDEP, Ge et al., 2018). Data can first be explored using heatmap, k-means clustering, and PCA. Differential expression analysis can be conducted through two different methods; limma and DESeq2 packages and several visualization plots are available (e.g., Venn diagrams, Volcano plots, genome maps). Based on Ensembl annotation, pathway enrichment can be performed from GO and KEGG annotation using several methods: GSEA, PAGE, GAGE or ReactomePA.

RNA-Seq data generated in this study over a two-week iron (Fe) starvation time-course

Wild-type *P. tricornutum* v 1.86 cells were grown in ESAW medium at 19°C (Dorrell, PNAS 2021), under 12hLight:12Dark cycle (50uE light). Fe-replete (Fe+) and Fe-depleted (Fe-)

ESAW media were both produced using iron-free reagents based on a protocol from Xia Gao (Gao et al., 2019). For the transcriptomic analyses, three different Fe limitation conditions were designed: three days Fe limitation as a short-term Fe- treatment (FeS), one week Fe limitation as a medium-term Fe- treatment (FeM) and two weeks Fe limitation as a long-term Fe- treatment (FeL). Fe-replete conditions (Fe+) were used as a control. RNA-Seq analysis were performed using two genetically distinct lines of *P. tricornutum* wild-type cells, transformed with pPhat and HA-Cas9 vectors without guide RNAs (Dorrell et al, 2024), to allow subsequent comparison of gene expression responses to mutants (data not shown). Reproductible transcriptome dynamics were observed for each culture, suggesting insertion of the pPhat and Cas9 vectors did not intrinsically bias Fe metabolism in these strains. Three biological replicates were performed for each cell line and condition. Cells were collected for the analyses at the exponential phase, and no other nutrient stress than Fe limitation influenced the results. Fast Fv/Fm (with 10% FP, 70% SP) tests were performed for all cell lines using a PAM (PAR-FluorPen FP 110, Photon Systems Instruments) prior to sampling. Fv/Fm values under Fe-replete values were measured at mean 0.62, suggesting that neither N nor Pi were growth-limiting in the media. Measured Fv/Fm values after 3 days Fe starvation were 0.55, suggesting Fe limitation of photosystem activity.

Total RNA was extracted from around 50 mg cell pellets by using the TRIzol reagent (T9424, Sigma-Aldrich) and according to (Dorrell et al, 2024). 24 DNAse-treated RNA libraries (4 conditions x 3 biological replica x 2 genetically distinct cell lines) were sequenced on a DNBseq Illumina platform (BGI Genomics Co., Ltd, Hongkong, China) with 100 bp paired-end sequencing. Raw reads were filtered by removing adaptor sequences, contamination and low-quality reads (reads containing over 40% bases with Q value < 20%) to obtain clean reads. Clean reads were mapped to the version 3 annotation of the *P. tricornutum* genome (Rastogi et al., 2018), and average TPM values for each gene in each library were calculated using DiatomicBase.

Author contributions

AF and CB conceived and supervised the project. EV coordinated the project, provided initial databases and drafted the manuscript. NZ designed the website, loaded the data with the technical help of PV. SL designed and performed RNA-Seq analysis under progressive Fe limitation. HCC, RGD, CD, RM and AF performed and interpreted case studies. KV, MF, HCC, RGD, AF and CB provided critical suggestions to the manuscript. All authors proofread and approved the manuscript.

Acknowledgements

The authors would like to thank the research community working on molecular aspects of diatoms for their feedback during the creation and curation of the website. NZ and EV thank Dr Ge and his team from the South Dakota State University for help in deploying iDEP on DiatOmicBase. NZ would like to thank Catherine Le Bihan, Maël Lefeuvre, Nolwenn Lavielle, Phi Phong Nguyen from the IBENS computing service for their technical support. We thank Mariella Ferrante, Svenja Mager, and Anna Santin for helping us to add *Pseudo-nitzschia multistriata* to DiatOmicBase. KV acknowledges Michiel Van Bel and Emmelien Vancaester for technical assistance and data curation within PLAZA Diatoms.

This work was supported principally by a grant from Gordon and Betty Moore Foundation GBMF8752. CB acknowledges additional support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Diatomic; grant agreement No. 835067), the French Government 'Investissements d'Avenir' programs MEMO LIFE (ANR-10-LABX-54) and PSL Research University (ANR-11-IDEX-0001-02). AF acknowledges funding from the Fondation Bettencourt-Schueller (Coups d'élan pour la recherche francaise-2018), the "Initiative d'Excellence" program (Grant "DYNAMO," ANR-11-LABX-0011-01) and EMBRC-FR – "Investments d'avenir" program (ANR-10-INBS-02). AF and CB acknowledges the ANR BrownCut (Project-ANR-19-CE20-0020) and Horizon Europe BlueRemediomics (grant agreement No. 101082304) projects. HCC acknowledges funding from ANR DiaLincs (ANR 19-CE43-0011-01). RGD acknowledges an ERC Starting Grant (grant number 101039760, "ChloroMosaic"). KV acknowledges Research Foundation-Flanders (FWO) for ELIXIR Belgium [I002819N] and BOF/GOA No. 01G01715 and 01G01323. MF acknowledges a Villum Young Investigator Grant (Villum Fonden grant number 37521).

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Data statement

Raw fastq data corresponding to new RNA-Seq performed under progressive Fe limitation is provided in NCBI BioProject number PRJNA936812.

Supporting information provided in a separate file:

Figure S1 : CPS III (Phatr3_J24195) genome browser capture displaying Phatr3 and Phatr2 annotations as well as peptides mapped using a proteogenomic pipeline. A zoom in the 5' region shows that two peptide sequences are mapped on an exon predicted by Phatr3 gene model but not by Phatr2. This gene model is also predicted by mapping of RNA-Seq reads (here from a study of McCarthy et al., 2017).

Figure S2 : A snapshot of the complete gene page for CPS III (Phatr3_J24195).

Figure S3 : Heatmap showing two-component regulator gene expression in light- and nutrient stress- related transcriptomes.

Figure S4 : Heatmap displaying transcription factor expression in light and nutrient stress related experiments.

Figure S5 : Schematic diagram of the culture regime used for Fe limitation experiments (BioProject number PRJNA936812).

References

Abbriano, R.M., George, J., Kahlke, T., Commault, A.S. and Fabris, M. (2023) Mobilization of a diatom mutator-like element () transposon inactivates the uridine monophosphate synthase (UMPS) locus in Phaeodactylum tricornutum. *The Plant Journal*, **115**, 926–936. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/tpj.16271 [Accessed July 17, 2024].

Agarwal, A., Di, R. and Falkowski, P.G. (2022) Light-harvesting complex gene regulation by a MYB-family transcription factor in the marine diatom, Phaeodactylum tricornutum. *Photosynth Res*, **153**, 59–70. Available at: https://doi.org/10.1007/s11120-022-00915-w [Accessed July 22, 2024].

Ait-Mohamed, O., Novák Vanclová, A.M.G., Joli, N., Liang, Y., Zhao, X., Genovesio, A., Tirichine, L., Bowler, C. and Dorrell, R.G. (2020) PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom Phaeodactylum tricornutum. *Frontiers in Plant Science*, **11**. Available at: https://www.frontiersin.org/article/10.3389/fpls.2020.590949 [Accessed February 28, 2022].

Allen, A.E., Dupont, C.L., Oborník, M., et al., (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*, **473**, 203–207. Available at: https://www.nature.com/articles/nature10074 [Accessed February 28, 2022].

Allen, A.E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P.J., Finazzi, G., Fernie, A.R. and Bowler, C. (2008) Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proceedings of the National Academy of Sciences*, **105**, 10438–10443. Available at: https://www.pnas.org/doi/full/10.1073/pnas.0711370105 [Accessed July 19, 2024].

Allen, J.F., Paula, W.B.M. de, Puthiyaveetil, S. and Nield, J. (2011) A structural phylogenetic map for chloroplast photosynthesis. *Trends in Plant Science*, **16**, 645–655. Available at: https://www.cell.com/trends/plant-science/abstract/S1360-1385(11)00226-3 [Accessed July 22, 2024].

Alverson, A.J., Jansen, R.K. and Theriot, E.C. (2007) Bridging the Rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Molecular Phylogenetics and Evolution*, **45**, 193–210. Available at: https://www.sciencedirect.com/science/article/pii/S1055790307001005 [Accessed October 5, 2023].

Annunziata, R., Ritter, A., Fortunato, A.E., et al., (2019) bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom Phaeodactylum tricornutum. *Proceedings of the National Academy of Sciences*, **116**, 13137–13142. Available at: https://www.pnas.org/doi/abs/10.1073/pnas.1819660116 [Accessed July 22, 2024].

Armbrust, E.V., Berges, J.A., Bowler, C., et al., (2004) The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism. *Science*, **306**, 79–86. Available at: https://www.science.org/doi/abs/10.1126/science.1101156 [Accessed October 28, 2021].

Ashburner, M., Ball, C.A., Blake, J.A., et al., (2000) Gene Ontology: tool for the unification of biology. *Nat Genet*, **25**, 25–29. Available at: https://www.nature.com/articles/ng0500_25 [Accessed February 28, 2022].

Ashworth, J., Turkarslan, S., Harris, M., Orellana, M.V. and Baliga, N.S. (2016) Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum. *Marine Genomics*, **26**, 21–28. Available at: https://www.sciencedirect.com/science/article/pii/S1874778715300441 [Accessed October 26, 2021].

Bailleul, B., Berne, N., Murik, O., et al., (2015) Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. *Nature*, **524**, 366–369. Available at: https://www.nature.com/articles/nature14599 [Accessed October 5, 2023].

Bailleul, B., Cardol, P., Breyton, C. and Finazzi, G. (2010) Electrochromism: a useful probe to study algal photosynthesis. *Photosynth Res*, **106**, 179–189.

Behnke, J. and LaRoche, J. (2020) Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). *European Journal of Phycology*, **55**, 339–360. Available at: https://doi.org/10.1080/09670262.2020.1744039 [Accessed July 22, 2024].

Blaby-Haas, C.E. and Merchant, S.S. (2019) Comparative and Functional Algal Genomics. *Annu Rev Plant Biol*, **70**, 605–638.

Bowler, C., Allen, A.E., Badger, J.H., et al., (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, **456**, 239–244. Available at: https://www.nature.com/articles/nature07410 [Accessed October 28, 2021].

Broddrick, J.T., Du, N., Smith, S.R., et al., (2019) Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom Phaeodactylum

tricornutum. *New Phytol*, **222**, 1364–1379. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6594073/ [Accessed July 22, 2024].

Buck, J.M., Sherman, J., Bártulos, C.R., et al., (2019) Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom Phaeodactylum tricornutum. *Nat Commun*, **10**, 4167. Available at: https://www.nature.com/articles/s41467-019-12043-6 [Accessed October 28, 2021].

Buels, R., Yao, E., Diesh, C.M., et al., (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biology*, **17**, 66. Available at: https://doi.org/10.1186/s13059-016-0924-1 [Accessed March 3, 2022].

Cohen, N.R., Mann, E., Stemple, B., Moreno, C.M., Rauschenberg, S., Jacquot, J.E., Sunda, W.G., Twining, B.S. and Marchetti, A. (2018) Iron storage capacities and associated ferritin gene expression among marine diatoms. *Limnology and Oceanography*, **63**, 1677–1691. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/lno.10800 [Accessed July 19, 2024].

Cruz de Carvalho, M.H., Sun, H.-X., Bowler, C. and Chua, N.-H. (2016) Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytologist*, **210**, 497–510. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.13787 [Accessed February 28, 2022].

Debit, A., Charton, F., Pierre-Elies, P., Bowler, C. and Cruz de Carvalho, H. (2023) Differential expression patterns of long noncoding RNAs in a pleiomorphic diatom and relation to hyposalinity. *Sci Rep*, **13**, 2440. Available at: https://www.nature.com/articles/s41598-023-29489-w [Accessed July 22, 2024].

Dell'Aquila, G. and Maier, U.G. (2020) Specific acclimations to phosphorus limitation in the marine diatom Phaeodactylum tricornutum. *Biological Chemistry*, **401**, 1495–1501. Available at: https://www.degruyter.com/document/doi/10.1515/hsz-2020-0197/html [Accessed February 28, 2022].

Delmont, T.O., Gaia, M., Hinsinger, D.D., et al., (2022) Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genomics*, **2**, 100123. Available at: https://www.sciencedirect.com/science/article/pii/S2666979X22000477 [Accessed May 17, 2023].

Dorrell, R.G., Zhang, Y., Liang, Y., et al., (2024) Complementary environmental analysis and functional characterization of lower glycolysis-gluconeogenesis in the diatom

plastid. *The Plant Cell*, koae168. Available at: https://doi.org/10.1093/plcell/koae168 [Accessed July 22, 2024].

Dyrløv Bendtsen, J., Nielsen, H., Heijne, G. von and Brunak, S. (2004) Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology*, **340**, 783–795. Available at: https://www.sciencedirect.com/science/article/pii/S0022283604005972 [Accessed June 2, 2023].

Falciatore, A., Jaubert, M., Bouly, J.-P., Bailleul, B. and Mock, T. (2020) Diatom Molecular Research Comes of Age: Model Species for Studying Phytoplankton Biology and Diversity[OPEN]. *The Plant Cell*, **32**, 547–572. Available at: https://doi.org/10.1105/tpc.19.00158 [Accessed October 28, 2021].

Falciatore, A. and Mock, T. eds. (2022) *The Molecular Life of Diatoms*, Cham: Springer International Publishing. Available at: https://link.springer.com/10.1007/978-3-030-92499-7 [Accessed July 19, 2024].

Ferrante, M.I., Broccoli, A. and Montresor, M. (2023) The pennate diatom Pseudo-nitzschia multistriata as a model for diatom life cycles, from the laboratory to the sea. *J Phycol*, **59**, 637–643.

Filloramo, G.V., Curtis, B.A., Blanche, E. and Archibald, J.M. (2021) Re-examination of two diatom reference genomes using long-read sequencing. *BMC Genomics*, **22**, 379. Available at: https://doi.org/10.1186/s12864-021-07666-3 [Accessed May 17, 2023].

Flori, S., Jouneau, P.-H., Bailleul, B., et al., (2017) Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nat Commun*, **8**, 15885. Available at: https://www.nature.com/articles/ncomms15885 [Accessed October 28, 2021].

Fukasawa, Y., Tsuji, J., Fu, S.-C., Tomii, K., Horton, P. and Imai, K. (2015) MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites. *Mol Cell Proteomics*, **14**, 1113–1126. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4390256/ [Accessed June 2, 2023].

Gao, X., Bowler, C. and Kazamia, E. (2021) Iron metabolism strategies in diatoms. *Journal of Experimental Botany*, **72**, 2165–2180. Available at: https://doi.org/10.1093/jxb/eraa575 [Accessed October 28, 2021].

Ge, S.X., Son, E.W. and Yao, R. (2018) iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*, **19**,

534. Available at: https://doi.org/10.1186/s12859-018-2486-6 [Accessed February 28, 2022].

Giguere, D.J., Bahcheli, A.T., Slattery, S.S., Patel, R.R., Browne, T.S., Flatley, M., Karas, B.J., Edgell, D.R. and Gloor, G.B. (2022) Telomere-to-telomere genome assembly of Phaeodactylum tricornutum. *PeerJ*, **10**, e13607.

Grouneva, I., Rokka, A. and Aro, E.-M. (2011) The Thylakoid Membrane Proteome of Two Marine Diatoms Outlines Both Diatom-Specific and Species-Specific Features of the Photosynthetic Machinery. *J. Proteome Res.*, **10**, 5338–5353. Available at: https://doi.org/10.1021/pr200600f [Accessed July 22, 2024].

Groussman, R.D., Parker, M.S. and Armbrust, E.V. (2015) Diversity and Evolutionary History of Iron Metabolism Genes in Diatoms. *PLOS ONE*, **10**, e0129081. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0129081 [Accessed July 22, 2024].

Gruber, A., Rocap, G., Kroth, P.G., Armbrust, E.V. and Mock, T. (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*, **81**, 519–528. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4329603/ [Accessed June 2, 2023].

Grypioti, E., Richard, H., Kryovrysanaki, N., Jaubert, M., Falciatore, A., Verret, F. and Kalantidis, K. (2024) Dicer-dependent heterochromatic small RNAs in the model diatom species Phaeodactylum tricornutum. *New Phytologist*, **241**, 811–826. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.19429 [Accessed July 22, 2024].

Gschloessl, B., Guermeur, Y. and Cock, J.M. (2008) HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*, **9**, 393. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2567999/ [Accessed June 2, 2023].

Gundermann, K., Schmidt, M., Weisheit, W., Mittag, M. and Büchel, C. (2013) Identification of several sub-populations in the pool of light harvesting proteins in the pennate diatom *Phaeodactylum tricornutum*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1827**, 303–310. Available at: https://www.sciencedirect.com/science/article/pii/S0005272812010742 [Accessed July 19, 2024].

Helliwell, K.E., Harrison, E.L., Christie-Oleza, J.A., et al., (2021) A Novel Ca2+ Signaling Pathway Coordinates Environmental Phosphorus Sensing and Nitrogen Metabolism in Marine Diatoms. *Current Biology*, **31**, 978-989.e4. Available at:

https://www.cell.com/current-biology/abstract/S0960-9822(20)31821-2 [Accessed July 19, 2024].

Hermann, D., Egue, F., Tastard, E., et al., (2014) An introduction to the vast world of transposable elements – what about the diatoms? *Diatom Research*, **29**, 91–104. Available at: https://doi.org/10.1080/0269249X.2013.877083 [Accessed October 28, 2021].

Heydarizadeh, P., Marchand, J., Chenais, B., Sabzalian, M.R., Zahedi, M., Moreau, B. and Schoefs, B. (2014) Functional investigations in diatoms need more than a transcriptomic approach. *Diatom Research*, **29**, 75–89. Available at: https://doi.org/10.1080/0269249X.2014.883727 [Accessed February 28, 2022].

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res*, **35**, W585–W587. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933216/ [Accessed June 2, 2023].

Jaubert, M., Duchêne, C., Kroth, P., Rogato, A., Bouly, J.-P. and Falciatore, A. (2022) Sensing and Signalling in Diatom Responses to Abiotic Cues. In pp. 607–639.

Johnson, G.N., Cardol, P., Minagawa, J. and Finazzi, G. (2014) Regulation of Electron Transport in Photosynthesis. In S. M. Theg and F.-A. Wollman, eds. *Plastid Biology*. New York, NY: Springer, pp. 437–464. Available at: https://doi.org/10.1007/978-1-4939-1136-3_16 [Accessed July 22, 2024].

Joshi-Deo, J., Schmidt, M., Gruber, A., Weisheit, W., Mittag, M., Kroth, P.G. and Büchel, C. (2010) Characterization of a trimeric light-harvesting complex in the diatom Phaeodactylum tricornutum built of FcpA and FcpE proteins. *Journal of Experimental Botany*, **61**, 3079–3087. Available at: https://doi.org/10.1093/jxb/erq136 [Accessed July 22, 2024].

Kanehisa, M. (2002) The KEGG database. *Novartis Found Symp*, **247**, 91–101; discussion 101-103, 119–128, 244–252.

Kazamia, E., Mach, J., McQuaid, J.B., et al., (2022) In vivo localization of iron starvation induced proteins under variable iron supplementation regimes in. *Plant Direct*, **6**, e472. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/pld3.472 [Accessed July 22, 2024].

Kazamia, E., Sutak, R., Paz-Yepes, J., et al., (2018) Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Science Advances*, **4**, eaar4536.

Available at: https://www.science.org/doi/full/10.1126/sciadv.aar4536 [Accessed July 22, 2024].

Keeling, P.J., Burki, F., Wilcox, H.M., et al., (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLOS Biology*, **12**, e1001889. Available at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001889 [Accessed February 28, 2022].

Kumar, M., Zuniga, C., Tibocha-Bonilla, J.D., Smith, S.R., Coker, J., Allen, A.E. and Zengler, K. (2022) Constraint-Based Modeling of Diatoms Metabolism and Quantitative Biology Approaches. In A. Falciatore and T. Mock, eds. *The Molecular Life of Diatoms*. Cham: Springer International Publishing, pp. 775–808. Available at: https://doi.org/10.1007/978-3-030-92499-7_26 [Accessed May 22, 2023].

Kwon, D.Y., Vuong, T.T., Choi, J., Lee, T.S., Um, J.-I., Koo, S.Y., Hwang, K.T. and Kim, S.M. (2021) Fucoxanthin biosynthesis has a positive correlation with the specific growth rate in the culture of microalga Phaeodactylum tricornutum. *J Appl Phycol*, **33**, 1473–1485. Available at: https://doi.org/10.1007/s10811-021-02376-5 [Accessed July 22, 2024].

Lampe, R.H., Mann, E.L., Cohen, N.R., Till, C.P., Thamatrakoln, K., Brzezinski, M.A., Bruland, K.W., Twining, B.S. and Marchetti, A. (2018) Different iron storage strategies among bloom-forming diatoms. *Proceedings of the National Academy of Sciences*, **115**, E12275–E12284. Available at: https://www.pnas.org/doi/full/10.1073/pnas.1805243115 [Accessed July 19, 2024].

Leister, D., Marino, G., Minagawa, J. and Dann, M. (2022) An ancient function of PGR5 in iron delivery? *Trends in Plant Science*, **27**, 971–980. Available at: https://www.cell.com/trends/plant-science/abstract/S1360-1385(22)00127-3 [Accessed July 22, 2024].

Lepetit, B., Campbell, D., Lavaud, J., Büchel, C., Goss, R. and Bailleul, B. (2022) Photosynthetic Light Reactions in Diatoms. II. The Dynamic Regulation of the Various Light Reactions. In Springer International Publishing, p. 423. Available at: https://hal.science/hal-03672201 [Accessed July 19, 2024].

Levitan, O., Dinamarca, J., Zelzion, E., et al., (2015) Remodeling of intermediate metabolism in the diatom Phaeodactylum tricornutum under nitrogen stress. *PNAS*, **112**,

412–417. Available at: https://www.pnas.org/content/112/2/412 [Accessed February 28, 2022].

Li, J., Zhang, K., Lin, X., Li, L. and Lin, S. (2022) Phytate as a Phosphorus Nutrient with Impacts on Iron Stress-Related Gene Expression for Phytoplankton: Insights from the Diatom Phaeodactylum tricornutum. *Applied and Environmental Microbiology*, **88**, e02097-21. Available at: https://journals.asm.org/doi/10.1128/aem.02097-21 [Accessed July 22, 2024].

Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Lodeyro, A.F., Ceccoli, R.D., Pierella Karlusich, J.J. and Carrillo, N. (2012) The importance of flavodoxin for environmental stress tolerance in photosynthetic microorganisms and transgenic plants. Mechanism, evolution and biotechnological potential. *FEBS Letters*, **586**, 2917–2924. Available at: https://www.sciencedirect.com/science/article/pii/S0014579312005947 [Accessed July 22, 2024].

Lommer, M., Specht, M., Roy, A.-S., et al., (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology*, **13**, R66. Available at: http://dx.doi.org/10.1186/gb-2012-13-7-r66 [Accessed June 26, 2016].

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550. Available at: https://doi.org/10.1186/s13059-014-0550-8 [Accessed February 28, 2022].

Madhuri, S., Lepetit, B., Fürst, A.H. and Kroth, P.G. (2024) A Knockout of the Photoreceptor Pt*Aureo1a* Results in Altered Diel Expression of Diatom Clock Components. *Plants*, **13**, 1465. Available at: https://www.mdpi.com/2223-7747/13/11/1465 [Accessed July 22, 2024].

Malviya, S., Scalco, E., Audic, S., et al., (2016) Insights into global diatom distribution and diversity in the world's ocean. *PNAS*, **113**, E1516–E1525. Available at: https://www.pnas.org/content/113/11/E1516 [Accessed October 26, 2021].

Mann, M., Serif, M., Wrobel, T., et al., (2020) The Aureochrome Photoreceptor Pt*AUREO1a* Is a Highly Effective Blue Light Switch in Diatoms. *iScience*, **23**. Available at: https://www.cell.com/iscience/abstract/S2589-0042(20)30927-5 [Accessed February 28, 2022].

Marchetti, A., Parker, M.S., Moccia, L.P., Lin, E.O., Arrieta, A.L., Ribalet, F., Murphy, M.E.P., Maldonado, M.T. and Armbrust, E.V. (2009) Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature*, **457**, 467–470. Available at: https://www.nature.com/articles/nature07539 [Accessed July 22, 2024].

Martino, A.D., Meichenin, A., Shi, J., Pan, K. and Bowler, C. (2007) Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions1. *Journal of Phycology*, **43**, 992–1009. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1529-8817.2007.00384.x [Accessed February 28, 2022].

Matthijs, M., Fabris, M., Broos, S., Vyverman, W. and Goossens, A. (2016) Profiling of the Early Nitrogen Stress Response in the Diatom Phaeodactylum tricornutum Reveals a Novel Family of RING-Domain Transcription Factors. *Plant Physiology*, **170**, 489–498. Available at: https://doi.org/10.1104/pp.15.01300 [Accessed February 28, 2022].

Matthijs, M., Fabris, M., Obata, T., Foubert, I., Franco-Zorrilla, J.M., Solano, R., Fernie, A.R., Vyverman, W. and Goossens, A. (2017) The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum. *EMBO J*, **36**, 1559–1576. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5452028/ [Accessed February 28, 2022].

Mattick, J.S. (2023) RNA out of the mist. *Trends in Genetics*, **39**, 187–207. Available at: https://www.cell.com/trends/genetics/abstract/S0168-9525(22)00288-8 [Accessed July 22, 2024].

McCarthy, J.K., Smith, S.R., McCrow, J.P., et al., (2017) Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom. *The Plant Cell*, **29**, 2047–2070. Available at: https://doi.org/10.1105/tpc.16.00910 [Accessed February 28, 2022].

McQuaid, J.B., Kustka, A.B., Oborník, M., et al., (2018) Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature*, **555**, 534–537. Available at: https://www.nature.com/articles/nature25982 [Accessed July 22, 2024].

Mock, T., Otillar, R.P., Strauss, J., et al., (2017) Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. *Nature*, **541**, 536–540. Available at: https://www.nature.com/articles/nature20803 [Accessed July 16, 2024].

Morrissey, J., Sutak, R., Paz-Yepes, J., et al., (2015) A Novel Protein, Ubiquitous in Marine Phytoplankton, Concentrates Iron at the Cell Surface and Facilitates Uptake.

Current Biology, **25**, 364–371. Available at: http://www.sciencedirect.com/science/article/pii/S0960982214015632 [Accessed July 5, 2016].

Nagao, R., Yokono, M., Ueno, Y., et al., (2021) Enhancement of excitation-energy quenching in fucoxanthin chlorophyll *a*/*c*-binding proteins isolated from a diatom *Phaeodactylum tricornutum* upon excess-light illumination. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1862**, 148350. Available at: https://www.sciencedirect.com/science/article/pii/S0005272820302000 [Accessed July 19, 2024].

Nef, C., Madoui, M.-A., Pelletier, É. and Bowler, C. (2022) Whole-genome scanning reveals environmental selection mechanisms that shape diversity in populations of the epipelagic diatom Chaetoceros. *PLOS Biology*, **20**, e3001893. Available at: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001893 [Accessed July 16, 2024].

Nemoto, M., Iwaki, S., Moriya, H., Monden, Y., Tamura, T., Inagaki, K., Mayama, S. and Obuse, K. (2020) Comparative Gene Analysis Focused on Silica Cell Wall Formation: Identification of Diatom-Specific SET Domain Protein Methyltransferases. *Mar Biotechnol*, **22**, 551–563. Available at: https://doi.org/10.1007/s10126-020-09976-1 [Accessed October 25, 2021].

Osuna-Cruz, C.M., Bilcke, G., Vancaester, E., et al., (2020) The Seminavis robusta genome provides insights into the evolutionary adaptations of benthic diatoms. *Nat Commun*, **11**, 3320.

Oudot-Le Secq, M.-P. and Green, B.R. (2011) Complex repeat structures and novel features in the mitochondrial genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana. *Gene*, **476**, 20–26. Available at: https://www.sciencedirect.com/science/article/pii/S0378111911000527 [Accessed October 28, 2021].

Oudot-Le Secq, M.-P., Grimwood, J., Shapiro, H., Armbrust, E.V., Bowler, C. and Green, B.R. (2007) Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage. *Molecular Genetics and Genomics*, **4**, 427–439. Available at: https://www.infona.pl//resource/bwmeta1.element.springer-55ad8850-d4ac-36e0-bac5-69f48e88b369 [Accessed October 26, 2021].

Patel, H., Ewels, P., Peltzer, A., et al., (2021) *nf-core/rnaseq: nf-core/rnaseq v3.5 - Copper Chameleon*, Zenodo. Available at: https://zenodo.org/record/5789421 [Accessed February 28, 2022].

Poulsen, N. and Kröger, N. (2023) Thalassiosira pseudonana (Cyclotella nana) (Hustedt) Hasle et Heimdal (Bacillariophyceae): A genetically tractable model organism for studying diatom biology, including biological silica formation. *Journal of Phycology*, **n/a**. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13362 [Accessed October 5, 2023].

Rastogi, A., Maheswari, U., Dorrell, R.G., et al., (2018) Integrative analysis of large scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. *Sci Rep*, **8**, 4834. Available at: https://www.nature.com/articles/s41598-018-23106-x [Accessed October 28, 2021].

Rastogi, A., Vieira, F.R.J., Deton-Cabanillas, A.-F., et al., (2020) A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom Phaeodactylum tricornutum. *ISME J*, **14**, 347–363. Available at: https://www.nature.com/articles/s41396-019-0528-3 [Accessed February 28, 2022].

Rogato, A., Amato, A., Iudicone, D., Chiurazzi, M., Ferrante, M.I. and Alcalà, M.R. d' (2015) The diatom molecular toolkit to handle nitrogen uptake. *Marine Genomics*, **24**, 95–108. Available at: https://www.sciencedirect.com/science/article/pii/S1874778715000951 [Accessed October 28, 2021].

Rogato, A., Richard, H., Sarazin, A., et al., (2014) The diversity of small non-coding RNAs in the diatom Phaeodactylum tricornutum. *BMC Genomics*, **15**, 698. Available at: https://doi.org/10.1186/1471-2164-15-698 [Accessed March 3, 2022].

Russo, M.T., Rogato, A., Jaubert, M., Karas, B.J. and Falciatore, A. (2023) Phaeodactylum tricornutum: An established model species for diatom molecular research and an emerging chassis for algal synthetic biology. *Journal of Phycology*, **59**, 1114–1122. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/jpy.13400 [Accessed July 19, 2024].

Sato, S., Nanjappa, D., Dorrell, R.G., et al., (2020) Genome-enabled phylogenetic and functional reconstruction of an araphid pennate diatom Plagiostriata sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. *Sci Rep*, **10**,

9449. Available at: https://www.nature.com/articles/s41598-020-65941-x [Accessed May 17, 2023].

Sétif, P. (2001) Ferredoxin and flavodoxin reduction by photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1507**, 161–179. Available at: https://www.sciencedirect.com/science/article/pii/S0005272801002055 [Accessed July 22, 2024].

Shen, C., Dupont, C.L. and Hopkinson, B.M. (2017) The diversity of CO2-concentrating mechanisms in marine diatoms as inferred from their genetic content. *Journal of Experimental Botany*, **68**, 3937–3948. Available at: https://doi.org/10.1093/jxb/erx163 [Accessed October 28, 2021].

Singer, D., Seppey, C.V.W., Lentendu, G., et al., (2021) Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. *Environment International*, **146**, 106262. Available at: https://www.sciencedirect.com/science/article/pii/S0160412020322170 [Accessed May 17, 2023].

Smith, S.R., Dupont, C.L., McCarthy, J.K., et al., (2019) Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun*, **10**, 4552. Available at: https://www.nature.com/articles/s41467-019-12407-y [Accessed February 28, 2022].

Smith, S.R., Gillard, J.T.F., Kustka, A.B., et al., (2016) Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLOS Genetics*, **12**, e1006490. Available at: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006490 [Accessed February 28, 2022].

Song, J., Zhao, H., Zhang, L., Li, Z., Han, J., Zhou, C., Xu, J., Li, X. and Yan, X. (2023) The Heat Shock Transcription Factor PtHSF1 Mediates Triacylglycerol and Fucoxanthin Synthesis by Regulating the Expression of GPAT3 and DXS in Phaeodactylum tricornutum. *Plant Cell Physiol*, **64**, 622–636.

Statello, L., Guo, C.-J., Chen, L.-L. and Huarte, M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol*, **22**, 96–118. Available at: https://www.nature.com/articles/s41580-020-00315-9 [Accessed July 22, 2024].

Tréguer, P.J., Sutton, J.N., Brzezinski, M., et al., (2021) Reviews and syntheses: The biogeochemical cycle of silicon in the modern ocean. *Biogeosciences*, **18**, 1269–1289. Available at: https://bg.copernicus.org/articles/18/1269/2021/ [Accessed October 28, 2021].

Turnšek, J., Brunson, J.K., Viedma, M. del P.M., Deerinck, T.J., Horák, A., Oborník, M., Bielinski, V.A. and Allen, A.E. (2021) Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway C. S. Hardtke, J. Kleine-Vehn, and C. Brownlee, eds. *eLife*, **10**, e52770. Available at: https://doi.org/10.7554/eLife.52770 [Accessed July 22, 2024].

Vancaester, E., Depuydt, T., Osuna-Cruz, C.M. and Vandepoele, K. (2020) Comprehensive and Functional Analysis of Horizontal Gene Transfer Events in Diatoms. *Molecular Biology and Evolution*, **37**, 3243–3257. Available at: https://doi.org/10.1093/molbev/msaa182 [Accessed July 16, 2024].

Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., Van de Peer, Y., Grimsley, N. and Piganeau, G. (2013) pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environmental Microbiology*, **15**, 2147–2153. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.12174 [Accessed February 28, 2022].

Vanormelingen, P., Verleyen, E. and Vyverman, W. (2009) The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. In W. Foissner and D. L. Hawksworth, eds. *Protist Diversity and Geographical Distribution*. Topics in Biodiversity and Conservation. Dordrecht: Springer Netherlands, pp. 159–171. Available at: https://doi.org/10.1007/978-90-481-2801-3_12 [Accessed October 28, 2021].

Veluchamy, A., Lin, X., Maumus, F., et al., (2013) Insights into the role of DNA methylation in diatoms by genome-wide profiling in Phaeodactylum tricornutum. *Nat Commun*, **4**, 2091. Available at: https://www.nature.com/articles/ncomms3091 [Accessed October 28, 2021].

Veluchamy, A., Rastogi, A., Lin, X., et al., (2015) An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biology*, **16**, 102. Available at: https://doi.org/10.1186/s13059-015-0671-8 [Accessed October 28, 2021].

Vernette, C., Lecubin, J., Sánchez, P., et al., (2022) The Ocean Gene Atlas v2.0: online exploration of the biogeography and phylogeny of plankton genes. *Nucleic Acids Research*,

50, W516–W526. Available at: https://doi.org/10.1093/nar/gkac420 [Accessed July 19, 2024].

Wu, Y., Chaumier, T., Manirakiza, E., Veluchamy, A. and Tirichine, L. (2023) PhaeoEpiView: an epigenome browser of the newly assembled genome of the model diatom Phaeodactylum tricornutum. *Sci Rep*, **13**, 8320. Available at: https://www.nature.com/articles/s41598-023-35403-1 [Accessed July 16, 2024].

Yang, M., Lin, X., Liu, X., Zhang, J. and Ge, F. (2018) Genome Annotation of a Model Diatom Phaeodactylum tricornutum Using an Integrated Proteogenomic Pipeline. *Molecular Plant*, **11**, 1292–1307. Available at: https://www.sciencedirect.com/science/article/pii/S1674205218302508 [Accessed February 28, 2022].

Zhao, X., Rastogi, A., Deton Cabanillas, A.F., et al., (2021) Genome wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell differentiation in Phaeodactylum tricornutum. *New Phytologist*, **229**, 3208–3220. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.17129 [Accessed May 17, 2023].

Zhou, L., Gao, S., Yang, W., Wu, S., Huan, L., Xie, X., Wang, X., Lin, S. and Wang, G. (2022) Transcriptomic and metabolic signatures of diatom plasticity to light fluctuations. *Plant Physiology*, **190**, 2295–2314. Available at: https://doi.org/10.1093/plphys/kiac455 [Accessed July 22, 2024].