METABOLIC REGULATION OF DIATOMS AND OTHER CHROMALVEOLATES

EDITED BY: Benoit Schoefs, Hanhua Hu, Justine Marchand and Kalina M. Manoylov

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88976-233-0
DOI 10.3389/978-2-88976-233-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

METABOLIC REGULATION OF DIATOMS AND OTHER CHROMALVEOLATES

Topic Editors:
Benoit Schoefs, Le Mans Université, France
Hanhua Hu, Institute of Hydrobiology, Chinese Academy of Sciences (CAS), China
Justine Marchand, Le Mans Université, France
Kalina M. Manoylov, Georgia College and State University, United States

Citation: Schoefs, B., Hu, H., Marchand, J., Manoylov, K. M., eds. (2022). Metabolic Regulation of Diatoms and Other Chromalveolates. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-233-0

Table of Contents

frontiers | Frontiers in Plant Science

EDITORIAL
published: 03 May 2022
doi: 10.3389/fpls.2022.897639

Editorial: Metabolic Regulation of Diatoms and Other Chromalveolates

Justine Marchand[1], *Hanhua Hu*[2], *Kalina Manoylov*[3] *and Benoît Schoefs*[1]*

[1] *Metabolism, Molecular Engineering of Microalgae and Applications, Laboratory Biologie des Organismes, Stress, Santé Environnement, IUML – FR 3473 CNRS, Le Mans University, Le Mans, France,* [2] *Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China,* [3] *Department of Biological and Environmental Sciences, Georgia College & State University, Milledgeville, GA, United States*

Keywords: microalgae, stress, omics, physiology, carbon metabolism

Editorial on the Research Topic

Metabolic Regulation of Diatoms and Other Chromalveolates

Microalgae are amongst the most abundant aquatic organisms. Microalgae form a polyphyletic group of organisms and diatoms belong to the Heterokonta clade (Kroth, 2015). This phylum evolved as a result of complex endosymbiosis and horizontal gene transfers from (cyano)bacteria and other microorganisms, including fungi [e.g., Thiriet-Rupert et al. (2016)], conferring them with unique biological features like efficient sequestering of dissolved CO_2, emitting a significant part of the oxygen (Benoiston et al., 2017), and performing efficient metabolic reorientation. Last but not the least, diatoms need silicon to build their cell wall by means of a network of nano-patterns forming very aesthetic decorations. To this end, diatoms rendered and still render enormous ecosystem services contributing significantly to several of the biogeochemical cycles and to the establishment of most of ocean food chains (Benoiston et al., 2017). Indeed, diatoms colonized successfully a wide range of environments, including the narrowest niches [e.g., Schoefs et al. (2020)] thanks to a very diversified and original metabolism [e.g., Allen et al. (2011)] and a high capacity to regulate it in order to acclimate to particular conditions [e.g., Heydarizadeh et al. (2017)]. The overload of these protective mechanisms results in cell death, making diatoms interesting organisms for the assessment of water quality (Szczepocka et al., 2021). In addition, microalgae have a huge potential for biotechnological applications (Sharma et al., 2021). However, biotechnology based on microalgae remains in its infancy and its development depends on the resolution of several bottlenecks (Vinayak et al., 2015) about which this theme takes stock:

- *A deeper knowledge of the basic cellular mechanisms*: Being photosynthetic organisms, diatoms convert sunlight energy into chemical energy used for running the Calvin-Benson-Basham (CBB) cycle along which CO_2 is fixed and converted into triose phosphates, ultimately used as building blocks for the synthesis of all the other cellular compounds. If the step succession of the CBB cycle is well-established, the regulation pathways at the transcriptional and post-transcriptional levels remain less clear. The article by Launay et al. takes stock of the different regulation levels (i.e., gene transcription, proteins production and enzyme activity). Interestingly, the redox regulation of the metabolic enzymes appears less important in diatoms than in green algae whereas the regulation at the transcriptional level seems to be widespread. The review also suggests that the role of post-translational modifications has been so far overlooked and needs further investigations. The contribution by Xie et al. on N-glycosylation in *Phaeodactylum tricornutum* Bohlin fills partially the gap. Using N-glycoproteomic and N-glycomic approaches, not less than 639 N-glycoproteins have been identified on the basis of 863 different N-glycopeptides.

OPEN ACCESS

Edited and reviewed by:
Miroslav Obornik,
Academy of Sciences of the Czechia Republic (ASCR), Czechia

***Correspondence:**
Benoît Schoefs
benoit.schoefs@univ-lemans.fr

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *16 March 2022*
Accepted: *28 March 2022*
Published: *03 May 2022*

Citation:
Marchand J, Hu H, Manoylov K and Schoefs B (2022) Editorial: Metabolic Regulation of Diatoms and Other Chromalveolates. Front. Plant Sci. 13:897639. doi: 10.3389/fpls.2022.897639

- To feed efficiently the CBB cycle with CO_2, diatoms import a considerable amount of CO_2 thanks to the carbon concentration mechanisms (CCMs). Two CCMs, namely the biophysical pathway and the biochemical pathway, have been recognized so far (Clement et al., 2017) but on the basis of the few taxa investigated [*P. tricornutum*: Kroth et al. (2008); *Thalassiosira pseudonana* (Hustedt) Hasle et Heimdal CCMP 1335: Kustka et al. (2014), Tanaka et al. (2014); *T. pseudonana* Hasle & Heim. strain CCAP 1085/12: Clement et al. (2016); *T. weissflogii* (Grunow) Fryxell et Hasle CCMP 1336 [current name *Conticribra weissflogii* (Grunow) Stachura-Such. & D.M. Williams]: Reinfelder et al. (2000, 2004); (Roberts et al., 2007)], it was concluded that only the biophysical pathway is commonly active in diatoms (Kroth, 2015). The molecular data generated along the Tara Oceans expeditions (Bork et al., 2015) allowed Pierella Karlusich et al. to extend this view to other diatoms. The triose phosphates generated along the CBB cycle are partly stored in storage polysaccharides with either α- or β-glucosidic linked glucan polymers, namely glycogen/starch or chrysolaminarins/paramylon, respectively. *In silico* analyses of genomics data allowed the identification of candidates coding new enzymes involved in storage polysaccharide biosynthetic pathways and the reconstitution of the evolutionary history of the distribution of these pathways in Stramenopiles (Chabi et al.).
- *In-depth knowledge of the mechanisms regulating the response to individual or combined stresses:* Living in a complex environment, like the ocean, is not easy because of the frequent, and often significant, variations of the environmental factors, which can have additive effects. Scarsini et al. used a multidisciplinary approach to investigate the metabolic reorientation induced by the transition from nitrogen-replete to nitrogen starvation conditions in the marine diatom *P. tricornutum* cultured in a turbidostat. The switch between the two equilibria is driven by the intracellular nitrogen availability and mostly involves intracellular carbon reutilization rather than *de novo* carbon fixation. Nevertheless, chloroplast is kept in a stand-by mode allowing a fast resuming upon nitrogen repletion. The reutilization of the carbon involves several catabolic pathways including that of branched amino acids (Pan et al., 2017). In this theme issue, Pan et al. compiled omics data for providing a broad view on the contribution of amino acids to TAG accumulation. In another publication of this theme issue, Thangaraj et al. studied the effects of stress combination (temperature and silica) on the marine diatom *Skeletonema dohrnii* Sarno & Kooistra. The study found evidence for specific mechanisms to cope with these conditions: at low temperature, carbon and cell lipid quotas were higher while phosphate assimilation was reduced. This contrasts with silicate-limited cells in which phosphate cell quota was high while that of nitrate was low. Proteins associated with carbon fixation and photorespiration were downregulated in both stress conditions, while the genes coding proteins involved in carbohydrate and lipid syntheses were upregulated, confirming that lipid accumulation in stressed diatoms constitutes a default response mechanism as proposed by Heydarizadeh et al. (2019).
- The biochemical and physiological responses to stress rely on modifications of the transcription patterns. The diversity of experimental conditions, including taxon, growth and stress conditions, although providing complementary data, often prevents the determination of common modules in the responses to different stresses. Ait-Mohamed et al. analyzed RNAseq datasets generated under varying stress using Weighted Gene Correlation Network Analysis and identified 28 modules of co-expressed genes that reveal the fundamental principles on which co-regulation of genes expression in *P. tricornutum* relies.
- *The obtention of efficient biomolecule production platforms:* Despite the recognition that microalgae, including diatoms, synthesize many molecules of interest [e.g., Mimouni et al. (2012)] and the availability of tools for the genetic improvement of certain taxa (George et al., 2020), only a handful of diatom taxa are used on an industrial scale for the production of biomolecules. As pointed out by Vinayak et al. (2015), the biotechnological processes based on microalgae would benefit from a deeper knowledge in the basic functioning of diatoms coupled to a wider use of the biodiversity. The article by Galas et al. and Chuberre et al. compares the main morphotypes of *P. tricornutum* from the cell organization and metabolism point of views. The studies reveal that despite a common cell organization the oval cells exhibit a unique metabolic signature and excrete proteins more rapidly than the other morphotypes, probably due to specific activation of the secretory machinery. This characteristic could be helpful for improving the efficiency of non-conventional downstream processes such as biocompatible extraction (Gateau et al., 2021).

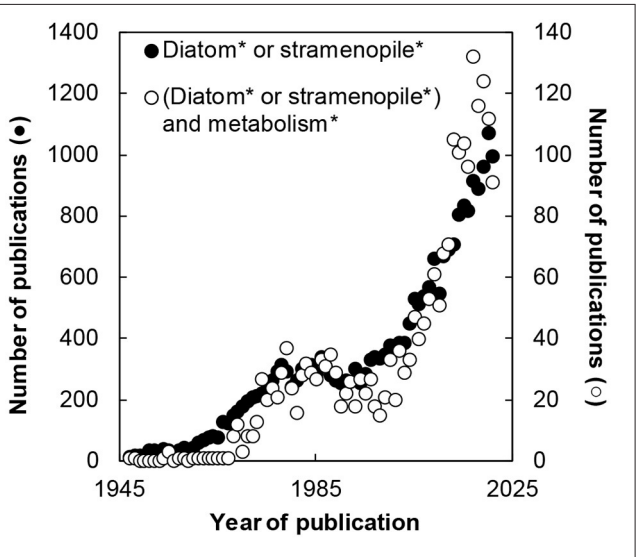


FIGURE 1 | Progression of the number of publications on diatom's metabolism.

- *Cost effective methods for biomolecule quantification:* The utilization of proxies, such as the optical density at 750 nm for the cell density, are often used to follow and characterize biological phenomena. Several proxies are based on extraction of biomolecules and a few are available for the characterization of cell processes and cellular quota. In this theme issue, Scarsini et al. established a Fourier Transform InfraRed (FTIR) microscopy method for the simultaneous quantification of lipids, carbohydrates, and proteins in diatoms. The limits of the method have been estimated and the means to circumvent them are proposed.

CONCLUSIONS

The research on diatoms and Stramenopiles is very dynamic. Since the World War II, more than 24,000 publications have been published with a title containing either word. *Circa* 10% of these articles are dedicated to their metabolism (**Figure 1**). This theme issue groups 11 articles describing the most recent research on the topic. These new data have been nicely welcomed by the scientific community with more than 30,000 views (https://www.frontiersin.org/research-topics/11978/metabolic-regulation-of-diatoms-and-other-chromalveolates - consulted on 2022 04 06) and 15 citations (WOS, all database, consulted on 2022 03 06).

AUTHOR CONTRIBUTIONS

KM, JM, HH, and BS contributed equally to the writing and editing of the editorial. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to thank Frontiers in Plant Sciences for recognizing the significance of our proposed topic and providing the help to the editors and reviewers. We hope that this issue will help reinforce the enthusiasm of all those interested in algal metabolisms. We are grateful for the valuable suggestions of the associate editor and reviewer.

REFERENCES

Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., Mccrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Benoiston, A.-S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., et al. (2017). The evolution of diatoms and their biogeochemical functions. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160397. doi: 10.1098/rstb.2016.0397

Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., and Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science* 348, 873–875. doi: 10.1126/science.aac5605

Clement, R., Dimnet, L., Maberly, S. C., and Gontero, B. (2016). The nature of the CO_2-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. *N. Phytol.* 209, 1417–1427. doi: 10.1111/nph.13728

Clement, R., Jensen, E., Prioretti, L., Maberly, S. C., and Gontero, B. (2017). Diversity of CO_2-concentrating mechanisms and responses to CO_2 concentration in marine and freshwater diatoms. *J. Exp. Bot.* 68, 3925–3935. doi: 10.1093/jxb/erx035

Gateau, H., Blanckaert, V., Veidl, B., Burlet-Schiltz, O., Pichereaux, C., Gargaros, A., et al. (2021). Application of pulsed electric fields for the biocompatible extraction of proteins from the microalga *Haematococcus pluvialis*. *Bioelectrochemistry* 137, 107588. doi: 10.1016/j.bioelechem.2020.107588

George, J., Kahlke, T., Abbriano, R. M., Kuzhiumparambil, U., Ralph, P. J., and Fabris, M. (2020). Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology. *Front. Bioeng. Biotechnol.* 8, 513. doi: 10.3389/fbioe.2020.00513

Heydarizadeh, P., Boureba, W., Zahedi, M., Huang, B., Moreau, B., Lukomska, E., et al. (2017). Response of CO_2-starved diatom *Phaeodactylum tricornutum* to light intensity transition. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160396. doi: 10.1098/rstb.2016.0396

Heydarizadeh, P., Veidl, B., Huang, B., Lukomska, E., Wielgosz-Collin, G., Couzinet-Mossion, A., et al. (2019). Carbon orientation in the diatom *Phaeodactylum tricornutum*: the effects of carbon limitation and photon flux density. *Front. Plant Sci.* 10, 471. doi: 10.3389/fpls.2019.00471

Kroth, P. G. (2015). The biodiversity of carbon assimilation. *J. Plant Physiol.* 172, 76–81. doi: 10.1016/j.jplph.2014.07.021

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jézéquel, V., Mock, T., Schnitzler Parker, M., et al. (2008). A model of carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* 3, e1426. doi: 10.1371/journal.pone.0001426

Kustka, A. B., Milligan, A. J., Zheng, H., New, A. M., Gates, C., Bidle, K. D., et al. (2014). Low CO_2 results in a rearrangement of carbon metabolism to support C_4 photosynthetic carbon assimilation in *Thalassiosira pseudonana*. *N. Phytol.* 204, 507–520. doi: 10.1111/nph.12926

Mimouni, V., Ulmann, L., Pasquet, V., Mathieu, M., Picot, L., Bougaran, G., et al. (2012). The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. *Curr. Pharmac. Biotechnol.* 13, 2733–2750. doi: 10.2174/138920112804724828

Pan, Y. F., Yang, J., Gong, Y. M., Li, X. L., and Hu, H. H. (2017). 3-hydroxyisobutyryl-CoA hydrolase involved in isoleucine catabolism regulates triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160409. doi: 10.1098/rstb.2016.0409

Reinfelder, J. R., Kraepiel, A. M. L., and Morel, F. M. M. (2000). Unicellular C_4 photosynthesis in a marine diatom. *Nature* 407, 996–999. doi: 10.1038/35039612

Reinfelder, J. R., Milligan, A. J., and Morel, F. M. M. (2004). The role of the C_4 pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiol.* 135, 2106–2111. doi: 10.1104/pp.104.041319

Roberts, K., Granum, E., Leegood, R. C., and Raven, J. A. (2007). C3 and C4 pathways of photosynthetic carbon assimilation in marine diatoms are under genetic, not environmental, control. *Plant Physiol.* 145, 230–235. doi: 10.1104/pp.107.102616

Schoefs, B., Van De Vijver, B., Wetzel, C., and Ector, L. (2020). From diatom species identification to ecological and biotechnological applications. *Bot. Lett.* 167, 2–6. doi: 10.1080/23818107.2020.1719883

Sharma, N., Simon, D. P., Diaz-Garza, A. M., Fantino, E., Messaabi, A., Meddeb-Mouelhi, F., et al. (2021). Diatoms biotechnology: various industrial applications for a greener tomorrow. *Front. Mar. Sci.* 8, 636613. doi: 10.3389/fmars.2021.636613

Szczepocka, E., Szulc, B., Szulc, K., Rakowska, B., and Zelazna-Wieczorek, J. (2021). Diatom indices in the biological assessment of the water quality based on the example of a small lowland river. *Oceanol. Hydrobiol. Stud.* 43, 265–273. doi: 10.2478/s13545-014-0141-z

Tanaka, R., Kikutani, S., Mahardika, A., and Matsuda, Y. (2014). Localization of enzymes relating to C_4 organic acid metabolisms in the marine diatom, *Thalassiosira pseudonana*. *Photosynth. Res.* 121, 251–263. doi: 10.1007/s11120-014-9968-9

Thiriet-Rupert, S., Carrier, G., Chénais, B., Trottier, C., Bougaran, G., Cadoret, J.-P., et al. (2016). Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genomics* 17, 282. doi: 10.1186/s12864-016-2610-9

Vinayak, V., Manoylov, K. M., Gateau, H., Blanckaert, V., Herault, J., Pencreac'h, G., et al. (2015). Diatom milking: a review and new approaches. *Mar. Drugs* 13, 2629–2665. doi: 10.3390/md13052629

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Marchand, Hu, Manoylov and Schoefs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

REVIEW
published: 16 July 2020
doi: 10.3389/fpls.2020.01033

Regulation of Carbon Metabolism by Environmental Conditions: A Perspective From Diatoms and Other Chromalveolates

Hélène Launay[1], *Wenmin Huang*[1,2], *Stephen C. Maberly*[3] *and Brigitte Gontero*[1]*

[1] BIP, Aix Marseille Univ CNRS, BIP UMR 7281, Marseille, France, [2] Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Center of Plant Ecology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, China, [3] UK Centre for Ecology & Hydrology, Lake Ecosystems Group, Lancaster Environment Centre, Lancaster, United Kingdom

Diatoms belong to a major, diverse and species-rich eukaryotic clade, the Heterokonta, within the polyphyletic chromalveolates. They evolved as a result of secondary endosymbiosis with one or more Plantae ancestors, but their precise evolutionary history is enigmatic. Nevertheless, this has conferred them with unique structural and biochemical properties that have allowed them to flourish in a wide range of different environments and cope with highly variable conditions. We review the effect of pH, light and dark, and CO_2 concentration on the regulation of carbon uptake and assimilation. We discuss the regulation of the Calvin-Benson-Bassham cycle, glycolysis, lipid synthesis, and carbohydrate synthesis at the level of gene transcripts (transcriptomics), proteins (proteomics) and enzyme activity. In contrast to Viridiplantae where redox regulation of metabolic enzymes is important, it appears to be less common in diatoms, based on the current evidence, but regulation at the transcriptional level seems to be widespread. The role of post-translational modifications such as phosphorylation, glutathionylation, etc., and of protein-protein interactions, has been overlooked and should be investigated further. Diatoms and other chromalveolates are understudied compared to the Viridiplantae, especially given their ecological importance, but we believe that the ever-growing number of sequenced genomes combined with proteomics, metabolomics, enzyme measurements, and the application of novel techniques will provide a better understanding of how this important group of algae maintain their productivity under changing conditions.

Keywords: Calvin cycle, CO_2 concentrating mechanism, *Phaeodactylum tricornutum*, redox regulation, *Thalassiosira pseudonana*

OPEN ACCESS

Edited by:
Benoit Schoefs,
Le Mans Université, France

Reviewed by:
Tore Brembu,
Norwegian University of Science and Technology, Norway
Pierre Cardol,
University of Liège, Belgium

****Correspondence:***
Brigitte Gontero
bmeunier@imm.cnrs.fr

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *29 April 2020*
Accepted: *23 June 2020*
Published: *16 July 2020*

Citation:
Launay H, Huang W, Maberly SC and Gontero B (2020) Regulation of Carbon Metabolism by Environmental Conditions: A Perspective From Diatoms and Other Chromalveolates. Front. Plant Sci. 11:1033. doi: 10.3389/fpls.2020.01033

INTRODUCTION

The chromalveolates are a polyphyletic eukaryote supergroup that includes many photosynthetic lineages including the cryptomonads, dinoflagellates, haptophytes, and heterokonts (also called stramenopiles) (Keeling, 2009). The phylogeny of diatoms and their allied groups is complicated (Dorrell et al., 2017; Falciatore et al., 2020) and while the chromalveolates are not now regarded as a

natural group we have retained the name here as a convenient and widely-used term. It has been estimated that over 50% of all formally described protists are chromalveolates (Cavalier-Smith, 2004; Cavalier-Smith and Chao, 2006). Within the diverse clade Heterokonta, diatoms (Bacillariophyceae) are photosynthetic microalgae that comprise between 30,000 and 100,000 species (Mann and Vanormelingen, 2013). They evolved about 250 Myrs ago (Medlin, 2016), are found today in all aquatic environments, and contribute about 20% to global primary production (Falkowski et al., 1998). Like other heterokonts, diatoms originated *via* serial endosymbioses (Stiller et al., 2014) and their chloroplasts derive from a red and a green algal endosymbiosis and also contain genes from prokaryotes, their eukaryotic host, and genes acquired by horizontal transfer (Moustafa et al., 2009; Deschamps and Moreira, 2012; Dorrell et al., 2017). Consequently, diatom genomes are enriched in genes from different origins and this combination has gifted them with unique metabolic features. In addition to the metabolism needed to produce a silica cell wall (Hildebrand et al., 2018) a functioning urea cycle is present (Allen et al., 2011; Nonoyama et al., 2019). Diatoms have an Entner-Doudoroff glycolytic pathway (Fabris et al., 2012) but lack the oxidative pentose phosphate (OPP) pathway in their chloroplast (Wilhelm et al., 2006; Kroth et al., 2008; Gruber et al., 2009). Their principal storage compound is a polysaccharide, chrysolaminarin (β-1,3 linked glucan) that is located in the vacuole rather than the chloroplast (Huang et al., 2018). Diatoms also have a large diversity of the metalloenzyme carbonic anhydrase (CA) that interconverts CO_2 and HCO_3^-. They possess seven of the eight known CA sub-classes, some of which can make use of metal cations other than the canonical zinc (Jensen et al., 2019a; Alissa et al., 2020; Morel et al., 2020). In diatoms, both the large and the small subunits of ribulose bisphosphate carboxylase-oxygenase (RuBisCO) are encoded by the chloroplast genome, in contrast to Viridiplantae where the small subunit is a nuclear encoded protein (Oudot-Le Secq et al., 2007). Moreover, most diatom plastid genomes, unlike those in Viridiplantae and the diatom *Seminavis robusta*, lack introns (Brembu et al., 2014). Also in diatoms, RuBisCO activation is mediated by the protein CbbX (Mueller-Cajar et al., 2011) that does not possess the cysteine residues found in RuBisCO activase (RCA) found in Viridiplantae, and thus cannot be redox regulated (Jensen et al., 2017). In addition, diatoms also have a pigment composition that substantially differs from plants (Carreto and Catoggio, 1976; Falkowski and Owens, 1980; Gilstad et al., 1993; Kuczynska et al., 2015). The most important accessory pigments in diatoms are fucoxanthin and chlorophyll *c* rather than chlorophyll *b* in Viridiplantae (Green, 2011). Also, like all photosynthetic eukaryotes and cyanobacteria, they contain xanthophylls that are derived from β carotene but in contrast, lack the α-carotene pathway. Diatoms are able to acclimate to a broad range of light irradiance and nutrient concentrations by adjusting their physiology and biochemical activity (Schoefs et al., 2017; Heydarizadeh et al., 2019). This requires a variety of mechanisms for balancing energy harvesting and light-energy consuming metabolic processes including carbon fixation (Wilhelm et al., 2006). In contrast to Viridiplantae, diatoms have a very low cyclic electron flow. To equilibrate the ratio of ATP to NADPH required for optimal photosynthesis, the chloroplast and the mitochondrion, that are physically in contact, exchange these compounds (Bailleul et al., 2015) (**Figure 1**).

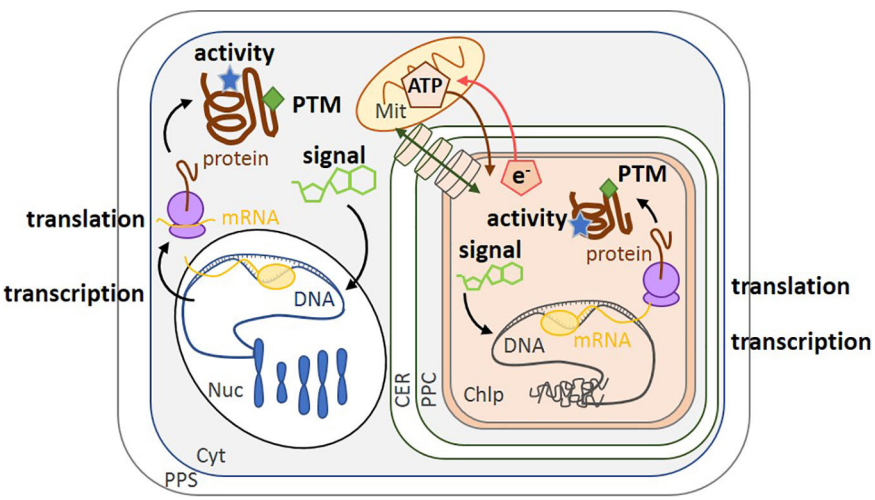


FIGURE 1 | Schematic of the different levels of regulation in diatoms. Transcription of the genome by mRNA polymerase (yellow circle) converts nuclear DNA (blue) and chloroplast DNA (gray) into mRNA (yellow). Translation by the ribosome (purple) converts mRNA into protein (brown). Regulation can also be modulated by enzyme activity (blue stars), post-translational modification (green diamonds) and small molecules such as co-factors or metabolites (green carbon skeletons). The transport of molecules between the cytoplasm and the chloroplast is represented by a green double-headed arrow across cylinders. ATP synthesized in the mitochondrion can be transported into the chloroplast (brown arrow), while the reducing power (e.g., NADPH, represented by e^-) of the chloroplast is transported to the mitochondrion (red arrow) (Bailleul et al., 2015). PPS, Periplasmic Space; CER, Chloroplast Endoplasmic Reticulum; PPC, Periplastidial Compartment; Mit, Mitochondrion; Chlp, Chloroplast; Cyt, Cytoplasm; PTM, Post-Translational Modification; Nuc, Nucleus.

Regulation can act on, and be studied at, a cascade of different levels from DNA (genomics), mRNA (transcriptomics), proteins (proteomics and post-translational modifications, PTMs), metabolites (metabolomics), and enzyme activity, because each approach provides different types of information (**Figure 1**). Genomes are powerful resources to determine if a specific gene is present while transcriptomics detect under what conditions it is expressed. Fully annotated diatom genome sequences are only available for *Thalassiosira pseudonana*, a marine centric diatom and *Phaeodactylum tricornutum*, a marine or coastal pennate diatom (Armbrust et al., 2004; Bowler et al., 2008). Other sequences are also available for *Fragilariopsis cylindrus* (Mock et al., 2017) and *Pseudo-nitzschia multiseries* (https://genome.jgi.doe.gov/portal/pages/tree-of-life.jsf) and there are further published genomes that are not yet publicly available (Tanaka et al., 2015; Traller et al., 2016; Basu et al., 2017; Villain et al., 2017; Ogura et al., 2018). Lauritano et al. (2019) reviewed the current development of omics approaches on microalgae. Of the 26 eukaryotic microalgal genomes they listed, 15 were from chromalveolates and of these, 8 were from diatoms. However, regulation should also be analyzed at the level of enzyme activity and/or metabolites (metabolomics), as these represent the final outcome of gene expression and activity (Prosser et al., 2014). The lifetime of an active enzyme, or of a metabolite, is related to its rate of synthesis and turnover. While the synthesis rates are on the order of 4–6 amino acids per second for enzymes (Stein and Frydman, 2019) and several seconds for metabolites (Nikolaev and Lohse, 2006), their turnover rates can vary from a few seconds to extended periods respectively. Enzyme activity is also modulated by PTMs, such as acetylation, phosphorylation, methylation, glycosylation and formation/dissociation of disulfide bonds. These types of modulation are very fast (rate on the order of few per second), reversible, and are the most flexible regulatory responses at the protein level (Prabakaran et al., 2012). In diatom RuBisCO, a number of post-translational modifications of the large subunit are present, including 4-hydroxyproline, β-hydroxyleucine, hydroxylated and nitrosylated cysteine, mono- and dihydroxylated lysine, and trimethylated lysine (Valegard et al., 2018). Nevertheless, in order to understand the full scope of regulation by post-translational modifications (Grabsztunowicz et al., 2017) in chromalveolates, more studies are needed on PTM and proteomics. Regulation of gene expression is itself dependent on earlier response regulators (for example, PTMs of histones and transcription factors) and on metabolite productions (for example, cAMP). As a consequence, upon environmental changes, regulation of gene expression occurs over a longer timescale of several minutes to hours (Chauton et al., 2013).

Relationships between mRNA level and protein expression can be observed though this might be influenced by biological (e.g., properties of mRNA and proteins, cell cycle status) and by technical problems (accurate quantification of these two biological molecules) (Maier et al., 2009; Ponnala et al., 2014). Therefore, there are discrepancies in the literature as regard to the extent of correlation between them. Net mRNA levels can be a major contributor to protein abundance, and for instance, positive relationship has been observed in yeast (Fournier et al., 2010), in the green alga *Chlamydomonas reinhardtii* (Castruita et al., 2011) and specifically in the diatom *T. pseudonana* (Clement et al., 2017b). Nevertheless, there are additional mechanisms, that control protein abundance including translational control and differential protein and mRNA degradation rates (Ponnala et al., 2014). However, since data on protein expression and activity are scarce, we have supplemented this type of information with data on gene regulation as a first step to assess how diatoms respond to environmental change, even though there is not always a direct and positive relationship between mRNA level, protein expression and finally, metabolic activity (**Figure 1**).

REGULATION OF PHOTOSYNTHESIS BY LIGHT AND DARK

Changing light levels affect many processes, including cell division, and diatoms can acclimate efficiently to light variation by altering the expression of different cell cycle genes such as cyclins and cyclin-dependent kinases genes (Huysman et al., 2013). Here, we focus on the best-studied effect of light, the regulation of photosynthesis, although there is much less information for diatoms than for the Viridiplantae (Jensen et al., 2017). In the Viridiplantae, that includes the Embryophyta, carbon fixation by the Calvin-Benson-Bassham (CBB) cycle is well known to be fine-tuned by dark-light transitions, involving regulation by pH (Werdan et al., 1975), Mg^{2+} (Portis and Heldt, 1976), metabolite concentration (Anderson, 1973; Pupillo and Giulianipiccari, 1975; Gardemann et al., 1983; Baalmann et al., 1994), and primarily by the redox state of key enzymes (Buchanan et al., 1980; Schurmann and Jacquot, 2000). Non-covalently bound "small molecules" or metabolites also affect the rates of redox-interconversion of each redox-regulated enzyme in Viridiplantae and this fine-tuning regulation is well-described in a review from Knuesting and Scheibe (2018).

Regulation by pH

In chloroplasts from Embryophyta, dark-to-light transitions are accompanied by a shift of the chloroplast internal pH from 7 in the dark to 8 in the light (Werdan and Heldt, 1973; Hauser et al., 1995). These changes directly regulate photosynthesis since many key chloroplastic enzymes have optimal activity at pH 8 and are much less active at pH 7 [reviewed in Gontero et al. (2007)]. In diatoms, pH responses have mainly been studied for external/environmental, rather than internal, pH. External pH can affect growth rate, silicon metabolism and biomineralization of *Conticribra weissflogii* (formerly known as *Thalassiosira weissflogii*) as well as its intracellular/cytoplasmic pH homeostasis (Herve et al., 2012). For other photosynthetic organisms, it is not the extracellular pH, but the intracellular pH in the chloroplast that is the critical factor for regulation of carbon acquisition, transport capacity and other metabolic processes. To our knowledge, internal pH has only been measured for a small number of diatoms. For *P. tricornutum* and *Cyclotella* sp. the pH was around 7 in the dark and 7.5 in the light (Colman and Rotatore, 1995), and for *Navicula pelliculosa*,

it was 7.4 in the dark and 7.6 in the light (Colman and Rotatore, 1988). We found no published values for the pH within diatom chloroplasts stroma. One of the few studies of the effect of pH on enzyme activity in chromalveolates is for the chloroplastic glyceraldehyde-3-phosphate dehydrogenase (GAPDH) that catalyzes the reversible reduction and dephosphorylation of 1,3-bisphosphoglycerate to produce glyceraldehyde-3-phosphate and inorganic phosphate. Avilan et al. (Avilan et al., 2012) compared the optimal pH of GAPDH in the freshwater diatom, *Asterionella formosa*, the freshwater eustigmatophyte, *Pseudocharaciopsis ovalis*, and the model green alga, *C. reinhardtii*. In *A. formosa*, GAPDH was still active at the pH occurring in the dark, assuming that the internal pH (pH 7) reflects the one in the chloroplast. This suggests that GAPDH is regulated by factors other than pH in this diatom, unlike in the green algal enzyme that was down-regulated at the pH that occurs in the dark. The response of GAPDH from the eustigmatophyte *P. ovalis* was similar to that of the green alga *C. reinhardtii*. We do not know the internal chloroplast pH for *P. ovalis* but if the dark-to-light pH transition in this species is similar to that of *C. reinhardtii*, GAPDH could be partly regulated by pH under dark-light transitions. The different regulation of GAPDH by pH in the two heterokonts, *A. formosa* and *P. ovalis*, might be the result of the diverse evolutionary history of chromalveolates. Another example of regulation by pH is the lumenal enzyme violaxanthin de-epoxidase [VDE, (Lavaud et al., 2012)] that is involved in dissipating excess light energy (Lohr and Wilhelm, 1999).

Beyond photosynthesis, the carbon metabolism of the marine diatom *Skeletonema costatum* is regulated by the pH of the growth medium (Taraldsvik and Myklestad, 2000). The content of the carbohydrate storage compound, chrysolaminarin (β-1,3 linked glucan) decreased from 7.1 $mg.L^{-1}$ at pH 6.5 to 0.2 $mg.L^{-1}$ at pH 9.4 and concomitantly, the total organic carbon as glucan also decreased from 60 to 10%. The total amino acid content also decreased from 7.41 to 2.51 $fmol.cell^{-1}$ when the pH of the growth medium increased (Taraldsvik and Myklestad, 2000). It is unclear if these are direct effects on carbon and nitrogen metabolism of external or internal pH or indirect effects linked to the greater external concentration of CO_2 at pH 6.5 than at pH 9.4. Nevertheless, to relate these physiological responses to enzyme activity regulation, the authors report results from a Norwegian PhD thesis (Kirkvold, 1994) that showed that the specific activity of glutamine synthetase, a key enzyme in the metabolic pathway of glutamine and glutamate synthesis, also decreased with increasing pH when measured *in vitro*.

Studies on the effect of pH on activity should be expanded to more enzymes and their optimal pH compared to the internal pH in dark and light in order to determine if enzyme activity is regulated by internal pH. The difficulty of working with enzymes from diatoms and from chromalveolates in general, is probably responsible for the lack of data for this important group. For instance, in order to extract proteins from diatoms, litres of culture are required and it is not always easy to measure activity. Expressing recombinant diatom enzymes in heterologous systems is also challenging with many enzymes found in the insoluble fractions (B. Gontero, personal communication). Measurement of internal pH is also an experimental tour de force. Colman and Rotatore used the 5,5-dimethyl-2,4-oxazolidinadione distribution method that distributes between the medium and the cell as a function of their respective pH (Colman and Rotatore, 1988; Colman and Rotatore, 1995). However, this method does not distinguish between the pH in the cytoplasm, chloroplast stroma or thylakoid lumen.

Regulation by the Redox State of Cysteine Residues

The redox control of enzyme activity in the Viridiplantae is primarily mediated by small proteins, thioredoxins, that are oxidized in the dark and reduced in the light (Buchanan et al., 1980; Buchanan, 2017). This regulation avoids futile cycles between the CBB and the OPP pathway, since both occur within the chloroplast, with enzymes from the CBB being active in the light and those from OPP being active in the dark. In contrast, in diatom plastids the OPP is incomplete, and presumably lacking (Kroth et al., 2008), and accordingly the regulation of their metabolism is different (Jensen et al., 2017). Moreover, diatoms have a high stromal reductant pressure and in contrast to Viridiplantae, metabolic activity in long dark periods leads to an enhanced reduction state of the plastoquinone pool. In the dark, since the plastoquinone pool is reduced, it may regulate redox-sensitive enzymes as is the case for algal nitrate reductase (Giordano et al., 2005). This avoids reducing equivalents to accumulate maintaining cellular redox poise (Wilhelm et al., 2006). In Viridiplantae in contrast, oxidizing conditions prevail in the dark, therefore suggesting that redox control may be different. Because of this unusual redox control, the redox regulation of diatom enzymes has been questioned (Wilhelm et al., 2006). However, diatoms possess many different thioredoxins, each encoded by a specific gene and located in different compartments, including the chloroplast. Most thioredoxins contain the regulatory cysteine residue in the conserved motif, WCGPC (Weber et al., 2009), thus they are likely to have specific regulation targets and some targets have been identified such as two CAs in *P. tricornutum* (Kikutani et al., 2012). The relatively few targets of thioredoxins currently identified in diatoms, contrasts with the 1188 targets found by combining qualitative and quantitative proteomic analyses in the *C. reinhardtii* thioredoxome (Perez-Perez et al., 2017).

Using a redox proteomics approach on *P. tricornutum*, Rosenwasser et al. identified the "redoxome", or in other words the redox-sensitive proteins, and demonstrated its involvement in photosynthesis, photorespiration, lipid biosynthesis, and nitrogen metabolism (Rosenwasser et al., 2014). In that case, however, the redox-sensitivity is a response to oxidative stress rather than light-dark transition even though reactive oxygen species are photo-induced, and increase in parallel to glutathione (GSH). GSH is a low-molecular-weight tripeptide that consists of cysteine (Cys), glutamic acid (Glu), and glycine (Gly) and is present in microorganisms, plants, and mammals (Zaffagnini et al., 2012). It can regulate protein activity by forming a mixed disulfide bridge between the thiol group of its Cys and an

accessible free thiol on a protein, a process known as protein *S*-glutathionylation (Zaffagnini et al., 2012; Marri et al., 2014). This post-translational modification can protect specific Cys residues from irreversible oxidation but can also modulate protein activities (Zaffagnini et al., 2012; Marri et al., 2014; Thieulin-Pardo et al., 2015). In *T. pseudonana*, a diurnal redox-related pattern has been observed in which GSH accumulates in the light, and decreases upon darkness, (Dupont et al., 2004) (**Figure 2**). However, direct regulation of the enzyme activities by glutathionylation in diatoms, or in other chromalveolates, in contrast to Viridiplantae, has not yet been studied, to the best of our knowledge.

Although the effect of glutathionylation on enzymes has not been studied in diatoms, the effect of other reducing agents such as dithiothreitol has been investigated, though understudied as compared to other photosynthetic organisms such as Cyanobacteria, Chlorophyta, Rhodophyta, and Embryophyta. The chloroplastic phosphoglycerate kinase belonging to the CBB cycle, catalyzes the ATP-Mg^{2+}-dependent phosphorylation of 3-phosphoglycerate (3-PGA) to 1,3-bisphosphoglycerate, in a reversible reaction and was redox-regulated in *P. tricornutum* (Belen Bosco et al., 2012). However, in our hands, PGK was not redox-regulated, as was also the case in *T. pseudonana*, in *Navicula pelliculosa* grown with sea water and fresh water medium and in a freshwater diatom, *A. formosa* (Jensen et al., 2019b). In contrast to the Viridiplantae, two enzymes that are unique to the CBB were not redox regulated (Michels et al., 2005; Maberly et al., 2010; Jensen et al., 2017). These include, sedoheptulose 1,7-bisphosphatase that irreversibly catalyzes the dephosphorylation of sedoheptulose-1,7-bisphosphate producing sedoheptulose-7-phosphate, and phosphoribulokinase (PRK) that irreversibly catalyzes the ATP-Mg^{2+}-dependent phosphorylation of ribulose-5-phosphate into ribulose-1,5-phosphate. The general lack or weak redox regulation of PRK in the chromalveolates (diatoms and other groups) seems to be related to its sequence, where the connectivity between two regulatory cysteine residues is crucial [at position 16 and 55 in *C. reinhardtii* (Maberly et al., 2010)]. In many photosynthetic organisms PRK can also be sequestered, and inactivated, in a PRK-GAPDH-CP12 complex, that has not yet been found in diatoms. The absence of the ternary complex with GAPDH in diatoms has been attributed to the absence of two cysteine residues on PRK (at position 243 and 249 numbered from the enzyme from *C. reinhardtii*) that are present in Cyanobacteria, Chlorophyta, Rhodophyta, and Embryophyta where the complex has been identified (Thieulin-Pardo et al., 2015). In contrast however, a ferredoxin-NADP reductase (FNR)-GAPDH-CP12

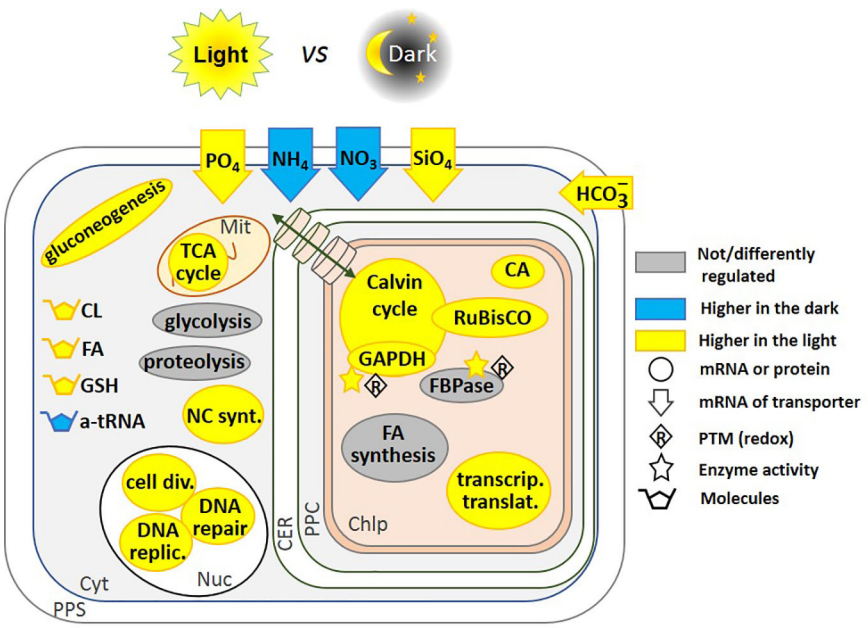


FIGURE 2 | Regulation of pathways by light and dark. This schematic includes regulatory pathways from *P. tricornutum* (Chauton et al., 2013; Bai et al., 2016), and *T. pseudonana* (Ashworth et al., 2013). Where up-regulation of the mRNA transcription preceded the start of the photoperiod, we have represented it as being higher in the light, this is the case for RuBisCO expression in *T. pseudonana* (Ashworth et al., 2013). Glycolysis and proteolysis pathways are up-regulated in the light in *T. pseudonana*, but down-regulated in *P. tricornutum*. In contrast, fatty acid synthesis is down-regulated in *T. pseudonana* but up-regulated in *P. tricornutum*. Fatty acid content increased after 3 h of light in *P. tricornutum* (Chauton et al., 2013), we have thus represented the molecule as being higher in the light. FBPase gene expression is not regulated, but its activity is redox-regulated (Michels et al., 2005; Mekhalfi et al., 2012). Redox regulation of enzyme activity has been shown on the isolated proteins *in vitro*, and by analogy with the Viridiplantae lineage, this could be related to light-to-dark transitions. The transport of molecules between the cytoplasm and the chloroplast is represented by a green double-headed arrow across cylinders. CL, Chrysolaminarin; FA, Fatty Acid; GSH, Glutathione; a-tRNA, aminoacyl-transfer RNA; CA, Carbonic Anhydrase; NC synt., Nucleotide synthesis; cell div., Cell division; DNA replic., DNA replication; PPS, Periplasmic Space; CER, Chloroplast Endoplasmic Reticulum; PPC, Periplastidial Compartment; Mit, Mitochondrion; Chlp, Chloroplast; Cyt, Cytoplasm; Nuc, Nucleus.

complex has been found in *A. formosa* (Mekhalfi et al., 2014). For the chloroplastic GAPDH, the regulation is more complex as discussed above in the pH regulation section, but in many diatoms, this enzyme seems to be redox regulated (Maberly et al., 2010; Mekhalfi et al., 2012; Mekhalfi et al., 2014; Jensen et al., 2019b).

Direct Light-Dark Control of Gene Expression

Regulation at the transcriptional level by light-dark transitions, occurs in Viridiplantae, and also in diatoms (Sun et al., 2003; Fey et al., 2005). In *T. pseudonana*, after 12 h of light, 1,859 genes were upregulated compared to cells exposed to 12 hours of dark, and inversely, after 12h of dark, 1,326 genes were up-regulated compared to cells exposed to 12 h of light (Ashworth et al., 2013). Among the most highly expressed genes after 12 h of light were the ones encoding enzymes for cell division, DNA replication and repair, carbon metabolism and oxidative phosphorylation while after 12 h of dark, the most highly expressed genes were those encoding ribosomal biogenesis, aminoacyl-tRNA and key photosynthetic enzymes (**Figure 2**). Some genes, such as that encoding RuBisCO, anticipates the diurnal cycle and is up-regulated before the onset of light. The dark-light expression pattern of genes was affected by growth phase (exponential *vs.* stationary). In the stationary phase, the expression of only a few genes fluctuated under dark-light transitions (Ashworth et al., 2013). One of these genes encodes a putative pyruvate carboxylase suggesting a switch toward other types of metabolism such as gluconeogenesis and lipid biosynthesis. This might be explained by the hypothesis raised by Norici et al. in *S. marinoi* (Norici et al., 2011) who postulated that the diatom re-routes its metabolism toward lipid biosynthesis, because of the relatively high volume-based energy content of lipids in an organism in which size decreases with vegetative cell divisions, thus requiring carbon allocation into more energy-compact compounds.

In the light, more than 4,500 transcripts were differentially expressed in *P. tricornutum*, including genes such as the one encoding for pyruvate transporter that had never been previously described in this organism (Chauton et al., 2013). This work shows that transcriptional regulation of carbohydrate and lipid metabolism occurs in diatoms (**Figure 2**). Indeed, the content of soluble glucans and lipids decreased in the dark and fatty acid biosynthesis genes were up-regulated within 30 min of a switch from dark to light. Fatty acid biosynthesis and the tricarboxylic acid (TCA) cycle are also tightly co-ordinated (Chauton et al., 2013). During the day, carbon skeletons are produced within the chloroplast while in the night these carbon-rich compounds are broken down in the mitochondria and the cytosol.

Interestingly, four carbon fixation enzymes were co-regulated in *P. tricornutum:* PGK, GAPDH, triose phosphate isomerase/GAPDH and PRK. Their mRNAs were all highest at the beginning of the light period (dawn) and lowest at the beginning of the dark period (dusk). Bai et al. (2016) showed however that the expression of PRK increased after 4 days of dark treatment using a proteomic approach. Similarly, the expression of phosphoribose isomerase, the enzyme that produces PRK substrate, as well as the expression of transaldolase and glyoxylase, increased after dark treatment.

Redox regulation of PRK is observed and well-studied in Viridiplantae, but in contrast in diatoms and other chromalveolates, as mentioned above, the activity of PRK was affected neither by oxidizing nor by reducing conditions (Maberly et al., 2010). Therefore, for enzymes that are not redox-regulated, regulation at the transcriptional level could allow a dark-light regulation of the CBB. The expression pattern of the gene of fructose-1,6-bisphosphatase (FBPase) that catalyzes the dephosphorylation of fructose-1,6-bisphosphate into fructose-6-phosphate and inorganic phosphate, was not induced to the same degree by light as other CBB enzymes. Since this enzyme can be redox regulated (Michels et al., 2005; Mekhalfi et al., 2012), its regulation could occur both at the level of activity and transcription.

Carbon acquisition is also stopped at night since in *P. tricornutum*, mRNA encoding for bicarbonate transporters of the Solute Carrier family 4 (SLC4) and for alpha-carbonic anhydrases, especially α-CA-VII, were much less abundant in the dark than in the light (Chauton et al., 2013). Recently, in the same organism, the pattern of mRNA levels at a photon irradiance of 30,300, 1,000 μmol photon m^{-2} s^{-1} differed at the lag, exponential and stationary phases of growth (Heydarizadeh et al., 2019).

Dark-treated *P. tricornutum* cells preferentially utilize carbon and nitrogen obtained from protein breakdown to increase lipid cell quotas at low cost (Bai et al., 2016). Long-term dark stress inhibited several key proteins involved in nitrogen assimilation and in the synthesis of the photosynthetic machinery. Simultaneously, key enzymes of glycolysis and the synthesis of fatty acids were induced apparently to assimilate the excess of C and N from protein breakdown. Uptake of other resources for growth are also light-regulated: transporters for uptake of phosphate and silica are higher in the light and nitrate and ammonium transporters are higher in the dark (Ashworth et al., 2013) (**Figure 2**). Data from the literature showing differential expression of proteins, transcripts and metabolites in light and dark are summarized in **Figure 2**.

Enzyme Activity, Metabolite Concentration, and Carbon Storage Compounds

Gene expression gives important clues on how light (quality, irradiance, duration) affects metabolism. However, it is also necessary to measure enzyme activity and metabolite concentration as these are the ultimate response to environmental change. For instance, an excess of light modifies lipid biosynthesis in the coastal marine diatom, *Skeletonema marinoi* (Norici et al., 2011). In *S. costatum*, carbohydrate increased with irradiance (Hitchcock, 1980) while lipids increased in *Chaetoceros calcitrans* (Harrison et al., 1990) and therefore the carbon allocation seems to be different and species-specific. Under different light-regimes, different species behave differently and the amount of essential fatty acids with growth irradiance is also species-specific. These examples illustrate that although enzyme activities have been

measured, the mechanism underlying the change in activity is unknown. For example, PEPCK that converts oxaloacetate into PEP and CO_2 in gluconeogenesis, increased 2.5-fold in cells of *S. marinoi* grown under low light (25 μmol.photon.m^{-2}.s^{-1}) *vs.* high light (250 μmol.photon.m^{-2}.s^{-1}). The authors suggested that this enzyme might be involved in the conversion of lipid to carbohydrates especially under low light (Norici et al., 2011). In contrast, since energy demand for lipid synthesis is much higher than for carbohydrate synthesis (Raven et al., 2005; Subramanian et al., 2013) under excess light, lipids represent a better sink for excess energy. As a consequence, lipid accumulation in high irradiance was observed in *S. marinoi*, although this is not always the case. In this species, other enzymes are probably not regulated by light since their activity remains unchanged. For instance, the activity of PEPC, that catalyzes the addition of bicarbonate (HCO_3^-) to PEP to produce oxaloacetate, was similar for cells grown at low or high light. This enzyme is involved in C4 metabolism and in anaplerotic reactions. Similarly, the activity of glutamine synthetase that is involved in photorespiration did not change in cells grown at low or high light.

EFFECT OF CO_2

Like other algae, diatoms exhibit a range of responses to varying CO_2 concentration, including effects on photophysiology, rate of photosynthesis and growth, chemical and pigment composition, and community species composition, but there are large species- and context-specific variations in the magnitude and sign of response (Boelen et al., 2011; Torstensson et al., 2012; Gao and Campbell, 2014; Dutkiewicz et al., 2015; Bach and Taucher, 2019; Jensen et al., 2019b). At the ocean surface, the air-equilibrium concentration of CO_2 (Dickson, 2010) varies between 5 and 25 μM depending on temperature (Raven, 1994; Tortell, 2000; Kim et al., 2006; Matsuda et al., 2011; Maberly and Gontero, 2017). This CO_2 concentration is insufficient to saturate the carboxylating enzyme, RuBisCO (Young et al., 2016) and may not saturate rates of diatom growth or photosynthesis (Riebesell et al., 1993; Dutkiewicz et al., 2015). This is mitigated by CO_2-concentrating mechanisms (CCMs) (Hopkinson et al., 2011) that are facultative and increase the concentration of CO_2 around RuBisCO, and are present in many phytoplankton. CCMs can involve biophysical or biochemical processes (Reinfelder et al., 2000; Reinfelder, 2011; Hopkinson et al., 2016) although the latter is controversial in diatoms (Clement et al., 2017a). The CCM regulation in diatoms is highly dependent on light as well as CO_2 concentration (Harada et al., 2006). However, the major determinant of the extent of CCM expression in *P. tricornutum* is CO_2 concentration, as it is in green algae (Matsuda and Kroth, 2014) and many components of CCM are suppressed under elevated CO_2 concentrations and induced at atmospheric levels or lower. The CO_2 concentration affects expression, and consequently the activity, of not only CCMs components but also that of enzymes from metabolic pathways such as the CBB cycle and glycolysis, though this is still understudied. Below, we compile information on the regulation by CO_2 at the transcriptional and/or the enzyme activity levels of enzymes involved in the CCM and other metabolic pathways.

CO_2-Concentrating Mechanisms (CCMs)

Biochemical CCM

In C4 metabolism, PEPC is the first carboxylating enzyme and traps bicarbonate into a C4 carbon compound. This compound is subsequently cleaved by a decarboxylase enzyme to provide a 3-carbon compound and CO_2, near the active site of RuBisCO (Sage, 2004). As mentioned above, the presence of C4 or biochemical CCM in diatoms does not seem to be universal. For instance, there is evidence for it in *C. weissflogii* (Reinfelder et al., 2000; Reinfelder, 2011; Hopkinson et al., 2016) but it is absent in *P. tricornutum* (Haimovich-Dayan et al., 2013; Clement et al., 2017a; Ewe et al., 2018). In the eustigmatophyte *Nannochloropsis oceanica*, a novel type of C4-based CCM was proposed to occur when cells were shifted from high CO_2 (50,000 ppm) to low CO_2 (100 ppm) (Wei et al., 2019). In this C4-based CCM, PEPC and PEPCK have been proposed to be involved in the primary inorganic carbon fixation steps in mitochondria, and not in chloroplasts. Subsequent decarboxylation of malate by a malic enzyme in the chloroplast enriches CO_2 in the vicinity of RuBisCO (**Figure 3**). Transcripts levels of some C4 enzymes were not altered by CO_2 concentration and do not suggest a classic C4 metabolism, but activities of PEPC and malic enzyme increased under low CO_2 (Wei et al., 2019). More work is required to confirm this interesting possibility.

Biophysical CCM

In the genomes of *P. tricornutum* and *T. pseudonana*, nine and thirteen CA gene sequences have been found respectively (Tachibana et al., 2011; Samukawa et al., 2014). In *P. tricornutum*, the two chloroplastic pyrenoidal β-CAs PtCA1 and 2, responded to CO_2 (Satoh et al., 2001; Harada and Matsuda, 2005; Harada et al., 2005) and later this was confirmed by Tachibana et al. by semi-quantitative RT-PCR (Tachibana et al., 2011). The activation of both PtCA1 and PtCA2 under CO_2 limitation involves three *cis*-regulatory elements, TGACGT, ACGTCA, and TGACGC, at a region minus 86 to minus 42 upstream of the transcription start site. These elements, CCRE1 to 3, are critical for the transcriptional response to ambient CO_2 *via* the level of the second messenger cAMP (Ohno et al., 2012; Tanaka et al., 2016). The sensing of CO_2 mediated by cAMP has been reported in cyanobacteria, fungi and mammals (Matsuda et al., 2011) and also in *T. pseudonana* (Hennon et al., 2015; Young and Morel, 2015). The transcriptional activation of PtCA2 in response to the decrease in CO_2 concentration was strongly light-dependent, such that either CO_2 or the absence of light can down-regulate the promoter. In contrast, CO_2 concentration and light have additive effects on the regulation of PtCA1 (Tanaka et al., 2016). It is worth remarking that both PtCA1 and PtCA2 were post translationally regulated by redox modifications *via* thioredoxins (Kikutani et al., 2012) (**Figure 3**).

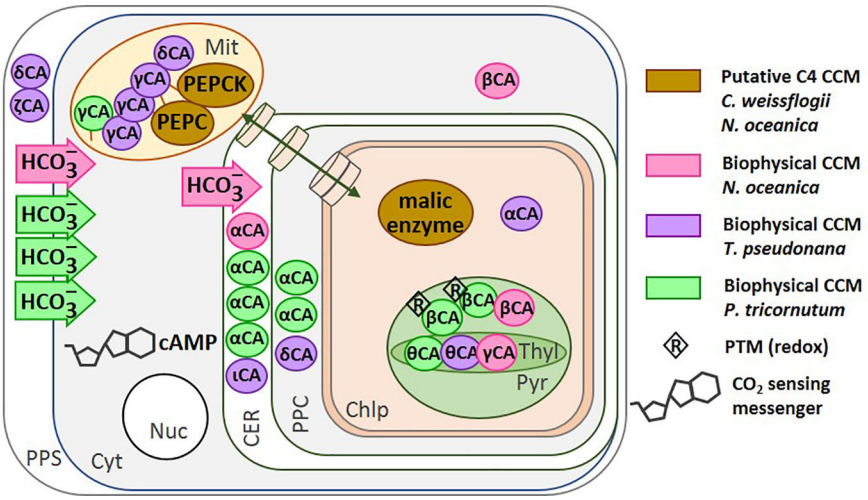


FIGURE 3 | Location of the CCM components in different diatom species. The carbonic anhydrases (circles) and bicarbonate transporters (arrows) are shown for: *N. oceanica* (Wei et al., 2019) (magenta), *T. pseudonana* (Samukawa et al., 2014; Jensen et al., 2020) (purple), and *P. tricornutum* (Jensen et al., 2020) (green). The putative C4-CCM components proposed for *N. oceanica* (Wei et al., 2019) and *C. weissflogii* (Reinfelder et al., 2000; Ohno et al., 2012; Hopkinson et al., 2016; Tanaka et al., 2016). The transport of molecules between the cytoplasm and the chloroplast is represented by a green double-headed arrow across cylinders. Redox regulation of CA activity has been shown on the isolated *P. tricornutum* β-CA *in vitro*. PEPC, Phosphoenolpyruvate Carboxylase; PEPCK, Phosphoenolpyruvate Carboxykinase; CA, Carbonic Anhydrase; PPS, Periplasmic Space; CER, Chloroplast Endoplasmic Reticulum; PPC, Periplastidial Compartment; Mit, Mitochondrion; Chlp, Chloroplast; Cyt, Cytoplasm; Nuc, Nucleus. Figure adapted from Jensen et al. (2020).

Tachibana et al. also showed that three putative CA genes in *T. pseudonana*, CA-1, 3, and 7 (α-CA, ζ-CA, and δ-CA, respectively) were induced by decreasing CO_2, and function in CO_2-limited environments (Tachibana et al., 2011). Similarly, in *C. weissflogii*, both CO_2 and HCO_3^- uptake increased in response to a CO_2 decrease and this was accompanied by an increase in both internal and external CA activity (Burkhardt et al., 2001). A recent proteomic study on *T. pseudonana* acclimated to low CO_2 (50 ppm) revealed a new uncharacterized protein, later called LCIP63, for "low-CO_2-inducible protein of 63 kDa" that was up-regulated under low CO_2 (50 ppm) or at atmospheric CO_2 (400 ppm) (Clement et al., 2017a) but down-regulated when nitrogen, phosphorus or silicon were limiting conditions (Lin et al., 2017; Chen et al., 2018). LCIP63 was up-regulated in *T. pseudonana* growing at 300 ppm *vs.* 1,000 ppm CO_2 (Valenzuela et al., 2018). Recently, this protein was identified as a new CA [designed as iota CA, (Jensen et al., 2019a)] that uses Mn^{2+} as a co-factor instead of the more common divalent cation Zn^{2+} (Tsuji et al., 2017; Dimario et al., 2018). The gene of iota CA is also present in bacterial genomes (Jensen et al., 2019a) and recently, the gene encoding this enzyme from the gram negative bacterium *Burkholderia territorii* was cloned and the purified recombinant protein showed a CA activity using Zn^{2+} instead of Mn^{2+} (Del Prete et al., 2020) indicating that the use of Mn^{2+} as a co-factor could be a feature of diatoms.

In *N. oceanica*, transcriptomic, proteomic and metabolomic data are available for cells grown at high CO_2 (50,000 ppm) and low CO_2 (100 ppm); three of the CA transcripts were up-regulated (β-CA-2, β-CA-4 and α-CA-5) under low CO_2 (Wei et al., 2019). In addition, the transcripts of bicarbonate transporters of the SLC4 family and several of the ABC transporter family were also more abundant at low CO_2, indicating an active biophysical CCM in this organism. Similarly, in *P. tricornutum*, beside the regulation of numerous CAs by CO_2, three out of seven SLC4 genes were induced by low CO_2 and were highly inhibited by the anion exchange inhibitor 4,4'-diisothiocyanatostilbene 2,2'-disulphonate (Nakajima et al., 2013). In *T. pseudonana*, genes homologous to those in *P. tricornutum* have been found (Matsuda et al., 2017); however, their functionality at the protein level has not yet been studied. Two chloroplast transporters of the bestrophin family of anion channels that are permeable to HCO_3^- were also more abundant at low CO_2 and may play a role in the biophysical CCM of this diatom (Kustka et al., 2014). These data are summarized in **Figure 3**.

Depending on future CO_2 emission scenarios, atmospheric levels of CO_2 are likely to reach 800 ppm by 2,100 (IPCC, 2014; Gattuso et al., 2015). This relates to a temperature-dependent dissolved CO_2 concentration at the ocean surface of 25 to 50 μM. At this concentration, cAMP plays a crucial role in down-regulating CCM genes in *T. pseudonana*, in particular those encoding the chloroplastic δCA3, some transporters and some involved in photorespiration (e.g., glycolate oxidase). These photorespiration genes and CCM genes interestingly, belong to a single CO_2-responsive cluster that shares the same upstream *cis*-regulatory sequences found in *P. tricornutum* that is also responsible for down-regulation of the β-CA gene upon increased CO_2 (Ohno et al., 2012). Similarly, genes involved in photosynthesis, the TCA cycle, oxidative phosphorylation and protein degradation were down-regulated, while in contrast other genes involved in signalling mechanisms were up-regulated at 800 ppm compared to 400 ppm CO_2 (Hennon

et al., 2015). Not all these genes contain the upstream regulatory region, though they were highly affected. This regulation is likely an indirect effect linked to the impact of high CO_2 on the genes of structural maintenance of chromosomes (SMC), transcription factors, and histones.

Sensing a change, either an increase or a decrease, in external CO_2 concentrations through cAMP seems to be a general rule of gene regulation rather than an exception even in diatoms (Young and Morel, 2015).

Enzymes From Carbon Metabolism

Many related metabolic pathways are affected by CO_2 in addition to CCM expression. Gamma CA and the NADH-ubiquinone oxidoreductase complex (C1) are associated in many organisms (Cardol, 2011) and in *N. oceanica*, at very low [CO_2] (100 ppm), the genes coding for these two mitochondrial enzymes were up-regulated (**Figure 4**). Since they can facilitate transport of CO_2 produced by the TCA cycle and photorespiration toward the chloroplast, in the form of bicarbonate, they contribute to what is called a basal CCM (Wei et al., 2019). At the enzyme activity level, *T. pseudonana* cells grown at high CO_2 concentration (20,000 ppm) *vs.* atmospheric (400 ppm) displayed higher NADPH-dependent GAPDH and FBPase activity (**Figure 4**) indicating that the CBB cycle could be affected even though PRK activity was unaltered (Clement et al., 2017b). The activity of pyruvate kinase, a glycolytic enzyme, was also strongly stimulated when cells were grown at 20,000 ppm (**Figure 4**) while two other glycolytic enzymes, NADH-dependent GAPDH and glucose-6-phosphate dehydrogenase (G6PDH) were unaffected (Clement et al., 2017b). In this diatom, a model summarizing the effect of inorganic carbon limitation, based both on activities and protein expression profiles has been elaborated (Clement et al., 2017b). This model shows the remodelling of metabolism with a diversion of energy and resources toward carbon metabolism at high CO_2 and toward carbon capture at low CO_2. An increase in δ-CA gene expression, to capture as much CO_2 as possible, was also observed by Kutska et al. (2014) in *T. pseudonana*. Metabolism enzymes, as well as enzymes responsible for pigment synthesis and indirectly in light capture (**Figure 4**), were less abundant, while enzymes involved in reactive oxygen species (ROS) defence increased in order to avoid accumulation of ROS that might occur when energy is in excess. The increased activity of pyruvate kinase (**Figure 4**), observed in *T. pseudonana* seems to be shared by other diatoms, as it was observed in many diatoms from freshwater to seawater species grown under high CO_2 (20,000 ppm) *vs.* atmospheric CO_2 (400 ppm) (Jensen et al., 2019b). In addition, a modification of gene expression has been described that allows synthesis of either PEP or pyruvate under carbon shortage, indicating that pyruvate is an important hub in these organisms (Heydarizadeh et al., 2017; Heydarizadeh et al., 2019) (**Figure 4**).

CONCLUSIONS AND FUTURE DIRECTIONS

Physiological and genomic data are available for the response of some chromalveolates, especially diatoms, to light and CO_2.

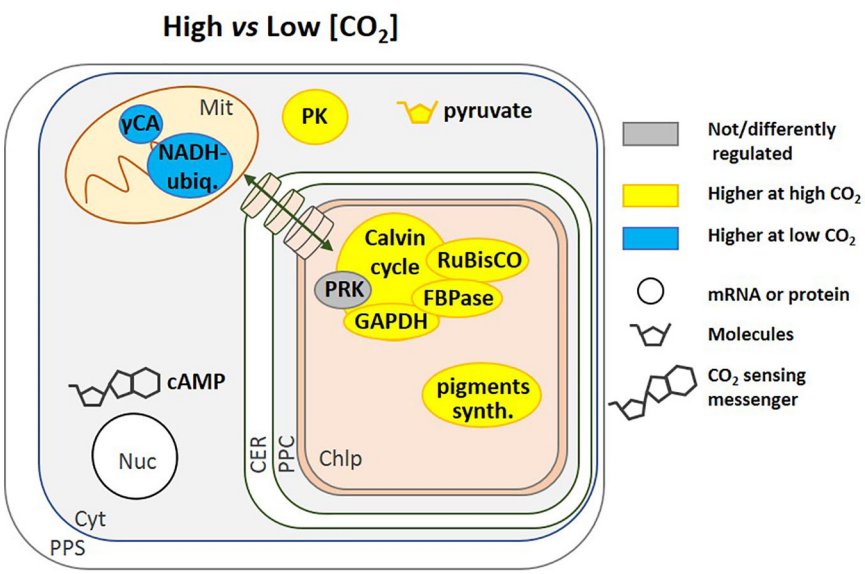


FIGURE 4 | Regulation of pathways by [CO_2]. This schematic includes regulatory pathways from different species, including *N. oceanica* (Wei et al., 2019) and *T. pseudonana* (Clement et al., 2017b; Jensen et al., 2019b). Cyclic AMP (cAMP) is a general [CO_2] signalling molecule that regulates gene expression (Ohno et al., 2012; Hennon et al., 2015; Young and Morel, 2015). High CO_2 corresponds to 20,000 to 50,000 ppm but in Hennon et al. (2015) it is 800 ppm and low CO_2 is 50 to 400 ppm. The transport of molecules between the cytoplasm and the chloroplast is represented by a green double-headed arrow across cylinders. NADH ubiq., NADH-ubiquinone oxidoreductase complex; PK, Pyruvate Kinase; pigments synth., pigments synthesis; CA, Carbonic Anhydrase; PPS, Periplasmic Space; CER, Chloroplast Endoplasmic Reticulum; PPC, Periplastidial Compartment; Mit, Mitochondrion; Chlp, Chloroplast; Cyt, Cytoplasm; Nuc, Nucleus.

They reveal the multitude and complexity of mechanisms that these organisms have evolved to cope with environmental variation. However, chromalveolates are still understudied compared to the Viridiplantae and more research is needed to unravel fully how this important group of algae maintain their productivity under changing conditions. There is a particular lack of information in diatoms on internal pH, especially in the chloroplast, on the identity of redox-regulated enzymes, on regulation by post-translational modification and on protein-protein interactions. There are existing and new methods that could be employed to tackle these knowledge gaps. For example, a range of pH-sensitive fluorescent probes are available to measure internal pH (Loiselle and Casey, 2010), although their low ability to penetrate the cell and organelle can be challenging. However, internal pH has also been measured successfully using the inorganic phosphate (^{31}P) nuclear magnetic resonance frequency in other organisms such as fungi (Hesse et al., 2000) and the anammox bacterium, *Kuenenia stuttgartiensis* (Van Der Star et al., 2010), and this could be applied to chromalveolates.

There are a growing number of studies taking advantage of proteomics to study diatom responses to stress, e.g. Muhseen et al. (2015) or Chen et al. (2018). Proteomics approaches have been used successfully in Viridiplantae to identify candidates for thioredoxin interactions (Montrichard et al., 2009; Perez-Perez et al., 2017). This could be extended to diatoms where there is a real challenge to assign specific targets to each of the many thioredoxins found in diatoms (Weber et al., 2009). Biotin-based proximity labelling approaches, coupled to quantitative proteomics, such as APEX BioID, are emerging tools for the study of protein-protein interactions (Santin et al., 2018; Beganton et al., 2019) that could be developed for chromalveolates. There is evidence for unusual PTMs involved in the regulation of RuBisCO from an arctic diatom (Valegard et al., 2018) and proteomics could also be a powerful approach to analyse these modifications. For instance, a phosphoproteomic analysis in *P. tricornutum* confirmed that phosphorylation occurs in many metabolic pathways (Chen et al., 2014).

Over-expressing or silencing a gene is starting to be applied to diatoms in order to determine their metabolic role (Hildebrand et al., 2017; Huang and Daboussi, 2017; Falciatore et al., 2020). Recently, tailored TALEN endonucleases and the CRISPR/Cas9 system have been utilized in diatoms (Daboussi et al., 2014; Hopes et al., 2016; Nymark et al., 2016), allowing knockout strains with targeted genetic modifications to be produced. An overview of the genetic toolbox currently available for performing stable genetic modifications in diatoms is reviewed in Kroth et al. and Falciatore et al. (Kroth et al., 2018; Falciatore et al., 2020).

It is clear that the techniques mentioned above in combination with genome sequencing, "omics" and targeted approaches, will allow the biology of diatoms and chromalveolates to be understood better. However, since many responses seem to be species-specific, a wider range of species need to be studied, especially those from non-marine systems, to produce a more complete picture of the functioning in this supergroup with a mosaic of multi-lineage genomes.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

CNRS and Aix Marseille are our current funders and the ANR-15-CE05-0021-03 allows us to work on diatom project.

ACKNOWLEDGMENTS

BG and HL were supported by the Centre National de la Recherche Scientifique (CNRS) and Aix-Marseille Université. BG is also supported by by the Agence Nationale de la Recherche (ANR-15-CE05-0021-03, SignauxBioNRJ). SM was supported by the UK Centre for Ecology & Hydrology. WH was supported by a fellowship for a visiting scholar from the Chinese Academy of Sciences.

REFERENCES

Alissa, S. A., Alghulikah, H. A., Alothman, Z. A., Osman, S. M., Del Prete, S., Capasso, C., et al. (2020). Inhibition survey with phenolic compounds against the delta- and eta-class carbonic anhydrases from the marine diatom *thalassiosira weissflogii* and protozoan *Plasmodium falciparum*. *J. Enzym. Inhib. Med. Ch.* 35, 377–382. doi: 10.1080/14756366.2019.1706089

Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., Mccrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Anderson, L. E. (1973). Regulation of pea leaf ribulose-5-phosphate kinase-activity. *Biochim. Biophys. Acta* 321, 484–488. doi: 10.1016/0005-2744(73)90190-3

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* 306, 79–86. doi: 10.1126/science.1101156

Ashworth, J., Coesel, S., Lee, A., Armbrust, E. V., Orellana, M. V., and Baliga, N. S. (2013). Genome-wide diel growth state transitions in the diatom *Thalassiosira pseudonana*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7518–7523. doi: 10.1073/pnas.1300962110

Avilan, L., Maberly, S. C., Mekhalfi, M., Plateau, J., Puppo, C., and Gontero, B. (2012). Regulation of glyceraldehyde-3-phosphate dehydrogenase in the eustigmatophyte *Pseudocharaciopsis ovalis* is intermediate between a chlorophyte and a diatom. *Eur. J. Phycol.* 47, 207–215. doi: 10.1080/09670262.2012.687459

Baalmann, E., Backhausen, J. E., Kitzmann, C., and Scheibe, R. (1994). Regulation of NADP-dependent glyceraldehyde-3-phosphate dehydrogenase-activity in spinach-chloroplasts. *Bot. Acta* 107, 313–320. doi: 10.1111/j.1438-8677.1994.tb00801.x

Bach, L. T., and Taucher, J. (2019). CO_2 effects on diatoms: a synthesis of more than a decade of ocean acidification experiments with natural communities. *Ocean Sci.* 15, 1159–1175. doi: 10.5194/os-15-1159-2019

Bai, X., Song, H., Lavoie, M., Zhu, K., Su, Y., Ye, H., et al. (2016). Proteomic analyses bring new insights into the effect of a dark stress on lipid biosynthesis in *Phaeodactylum tricornutum*. *Sci. Rep.* 6, 25494. doi: 10.1038/srep25494

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., Tanaka, A., et al. (2015). Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* 524, 366–369. doi: 10.1038/nature14599

Basu, S., Patil, S., Mapleson, D., Russo, M. T., Vitale, L., Fevola, C., et al. (2017). Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.* 215, 140–156. doi: 10.1111/nph.14557

Beganton, B., Coyaud, E., Mange, A., and Solassol, J. (2019). New approaches for protein-protein interaction study. *Med. Sci.* 35, 223–231. doi: 10.1051/medsci/2019035

Belen Bosco, M., Cristina Aleanzi, M., and Alvaro Iglesias, A. (2012). Plastidic phosphoglycerate kinase from *Phaeodactylum tricornutum*: on the critical role of cysteine residues for the enzyme function. *Protist* 163, 188–203. doi: 10.1016/j.protis.2011.07.001

Boelen, P., De Poll, W. H. V., Van Der Strate, H. J., Neven, I. A., Beardall, J., and Buma, A. G. J. (2011). Neither elevated nor reduced CO_2 affects the photophysiological performance of the marine Antarctic diatom *Chaetoceros brevis*. *J. Exp. Mar. Biol. Ecol.* 406, 38–45. doi: 10.1016/j.jembe.2011.06.012

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Brembu, T., Winge, P., Tooming-Klunderud, A., Nederbragt, A. J., Jakobsen, K. S., and Bones, A. M. (2014). The chloroplast genome of the diatom *Seminavis robusta*: new features introduced through multiple mechanisms of horizontal gene transfer. *Mar. Genom.* 16, 17–27. doi: 10.1016/j.margen.2013.12.002

Buchanan, B. B. (1980). Role of light in the regulation of chloroplast nzymes. *Ann. Rev. Plant Physiol.* 31, 341–374. doi: 10.1146/annurev.pp.31.060180.002013

Buchanan, B. B. (2017). The path to thioredoxin and redox regulation beyond chloroplasts. *Plant Cell Physiol.* 58, 1826–1832. doi: 10.1093/pcp/pcx119

Burkhardt, S., Amoroso, G., Riebesell, U., and Sultemeyer, D. (2001). CO_2 and HCO_3^- uptake in marine diatoms acclimated to different CO_2 concentrations. *Limnol. Oceanogr.* 46, 1378–1391. doi: 10.4319/lo.2001.46.6.1378

Cardol, P. (2011). Mitochondrial NADH:ubiquinone oxidoreductase (complex I) in eukaryotes: a highly conserved subunit composition highlighted by mining of protein databases. *Biochim. Biophys. Acta* 1807, 1390–1397. doi: 10.1016/j.bbabio.2011.06.015

Carreto, J. I., and Catoggio, J. A. (1976). Variations in pigment contents of the diatom *Phaeodactylum tricornutum* during growth. *Mar. Biol.* 36, 105–112. doi: 10.1007/BF00388433

Castruita, M., Casero, D., Karpowicz, S. J., Kropat, J., Vieler, A., Hsieh, S. I., et al. (2011). Systems biology approach in *Chlamydomonas* reveals connections between copper nutrition and multiple metabolic steps. *Plant Cell* 23, 1273–1292. doi: 10.1105/tpc.111.084400

Cavalier-Smith, T., and Chao, E. E. Y. (2006). Phylogeny and megasystematics of phagotrophic heterokonts (kingdom Chromista). *J. Mol. Evol.* 62, 388–420. doi: 10.1007/s00239-004-0353-8

Cavalier-Smith, T. (2004). "Chromalveolate diversity and cell megaevolution: interplay of membranes, genomes and cytoskeleton," in *Organelles, genomes and eukaryote phylogeny: an evolutionary synthesis in the age of genomics*. Eds. R. P. Hirt and D. S. Horner (London, UK: CRC Press LLC), 75–108.

Chauton, M. S., Winge, P., Brembu, T., Vadstein, O., and Bones, A. M. (2013). Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. *Plant Physiol.* 161, 1034–1048. doi: 10.1104/pp.112.206177

Chen, Z., Yang, M. K., Li, C. Y., Wang, Y., Zhang, J., Wang, D. B., et al. (2014). Phosphoproteomic analysis provides novel insights into stress responses in *Phaeodactylum tricornutum*, a model diatom. *J. Proteome. Res.* 13, 2511–2523. doi: 10.1021/pr401290u

Chen, X.-H., Li, Y.-Y., Zhang, H., Liu, J.-L., Xie, Z.-X., Lin, L., et al. (2018). Quantitative proteomics reveals common and specific responses of a marine diatom *Thalassiosira pseudonana* to different macronutrient deficiencies. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.02761

Clement, R., Jensen, E., Prioretti, L., Maberly, S. C., and Gontero, B. (2017a). Diversity of CO_2-concentrating mechanisms and responses to CO_2 concentration in marine and freshwater diatoms. *J. Exp. Bot.* 68, 3925–3935. doi: 10.1093/jxb/erx035

Clement, R., Lignon, S., Mansuelle, P., Jensen, E., Pophillat, M., Lebrun, R., et al. (2017b). Responses of the marine diatom *Thalassiosira pseudonana* to changes in CO_2 concentration: a proteomic approach. *Sci. Rep.* 7. doi: 10.1038/srep42333

Colman, B., and Rotatore, C. (1988). Uptake and accumulation of inorganic carbon by a fresh-water diatom. *J. Exp. Bot.* 39, 1025–1032. doi: 10.1093/jxb/39.8.1025

Colman, B., and Rotatore, C. (1995). Photosynthetic inorganic carbon uptake and accumulation in two marine diatoms. *Plant Cell Environ.* 18, 919–924. doi: 10.1111/j.1365-3040.1995.tb00601.x

Daboussi, F., Leduc, S., Marechal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* 5, 3831. doi: 10.1038/ncomms4831

Del Prete, S., Nocentini, A., Supuran, C. T., and Capasso, C. (2020). Bacterial iota-carbonic anhydrase: a new active class of carbonic anhydrase identified in the genome of the Gram-negative bacterium *Burkholderia territorii*. *J. Enzyme Inhib. Med. Chem.* 35, 1060–1068. doi: 10.1080/14756366.2020.1755852

Deschamps, P., and Moreira, D. (2012). Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol.* 4, 795–800. doi: 10.1093/gbe/evs053

Dickson, A. G. (2010). "The carbon dioxide system in seawater: equilibrium chemistry and measurements," in *Guide to best practices for ocean acidification research and data reporting*. Eds. U. Riebesell, V. J. Fabry, L. Hansson and J. P. Gattuso (Luxembourg: Publications Office of the European Union), 17–40.

Dimario, R. J., Machingura, M. C., Waldrop, G. L., and Moroney, J. V. (2018). The many types of carbonic anhydrases in photosynthetic organisms. *Plant Sci.* 268, 11–17. doi: 10.1016/j.plantsci.2017.12.002

Dorrell, R. G., Gile, G., Mccallum, G., Meheust, R., Bapteste, E. P., Klinger, C. M., et al. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife* 6, e23717. doi: 10.7554/eLife.23717

Dupont, C. L., Goepfert, T. J., Lo, P., Wei, L. P., and Ahnerz, B. A. (2004). Diurnal cycling of glutathione in marine phytoplankton: field and culture studies. *Limnol. Oceanogr.* 49, 991–996. doi: 10.4319/lo.2004.49.4.0991

Dutkiewicz, S., Morris, J. J., Follows, M. J., Scott, J., Levitan, O., Dyhrman, S. T., et al. (2015). Impact of ocean acidification on the structure of future phytoplankton communities. *Nat. Clim. Change* 5, 1002–1006. doi: 10.1038/nclimate2722

Ewe, D., Tachibana, M., Kikutani, S., Gruber, A., Bartulos, C. R., Konert, G., et al. (2018). The intracellular distribution of inorganic carbon fixing enzymes does not support the presence of a C_4 pathway in the diatom *Phaeodactylum tricornutum*. *Photosynth. Res.* 137, 263–280. doi: 10.1007/s11120-018-0500-5

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., and Baart, G. J. E. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.* 70, 1004–1014. doi: 10.1111/j.1365-313X.2012.04941.x

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. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158

Falkowski, P. G., and Owens, T. G. (1980). Light-shade adaptation - 2 strategies in marine-phytoplankton. *Plant Physiol.* 66, 592–595. doi: 10.1104/pp.66.4.592

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* 281, 200–206. doi: 10.1126/science.281.5374.200

Fey, V., Wagner, R., Brautigam, K., and Pfannschmidt, T. (2005). Photosynthetic redox control of nuclear gene expression. *J. Exp. Bot.* 56, 1491–1498. doi: 10.1093/jxb/eri180

Fournier, M. L., Paulson, A., Pavelka, N., Mosley, A. L., Gaudenz, K., Bradford, W. D., et al. (2010). Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. *Mol. Cell. Proteomics* 9, 271–284. doi: 10.1074/mcp.M900415-MCP200

Gao, K., and Campbell, D. A. (2014). Photophysiological responses of marine diatoms to elevated CO_2 and decreased pH: a review. *Funct. Plant Biol.* 41, 449–459. doi: 10.1071/FP13247

Gardemann, A., Stitt, M., and Heldt, H. W. (1983). Control of CO_2 fixation. Regulation of spinach ribulose-5-phosphate kinase by stromal metabolite levels. *BBA-Bioenergetics* 722, 51–60. doi: 10.1016/0005-2728(83)90156-1

Gattuso, J. P., Magnan, A., Billé, R., Cheung, W. W. L., Howes, E. L., Joos, F., et al. (2015). Contrasting futures for ocean and society from different anthropogenic CO_2 emissions scenarios. *Science* 349, aac4722. doi: 10.1126/science.aac4722 780

Gilstad, M., Johnsen, G., and Sakshaug, E. (1993). Photosynthetic parameters, pigment composition and respiration rates of the marine diatom *Skeletonema-costatum* grown in continuous light and a 12/12 h light-dark cycle. *J. Plankton Res.* 15, 939–951. doi: 10.1093/plankt/15.8.939

Giordano, M., Chen, Y. B., Koblizek, M., and Falkowski, P. G. (2005). Regulation of nitrate reductase in *Chlamydomonas reinhardtii* by the redox state of the plastoquinone pool. *Eur. J. Phycol.* 40, 345–352. doi: 10.1080/09670260500334263

Gontero, B., Avilan, L., and Lebreton, S. (2007). "Control of carbon fixation in chloroplasts," in *Annual plant reviews*. Eds. W. C. Plaxton and M. T. Mcmanus (Oxford: Blackwell Publishing), 187–218. doi: 10.1002/9780470988640.ch7

Grabsztunowicz, M., Koskela, M. M., and Mulo, P. (2017). Post-translational modifications in regulation of chloroplast function: recent advances. *Front. Plant Sci.* 8, 1–12. doi: 10.3389/fpls.2017.00240

Green, B. R. (2011). After the primary endosymbiosis: an update on the chromalveolate hypothesis and the origins of algae with Chl *c*. *Photosynth. Res.* 107, 103–115. doi: 10.1007/s11120-010-9584-2

Gruber, A., Weber, T., Bártulos, C. R., Vugrinec, S., and Kroth, P. G. (2009). Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *J. Basic Microb.* 49, 58–72. doi: 10.1002/jobm.200800339

Haimovich-Dayan, M., Garfinkel, N., Ewe, D., Marcus, Y., Gruber, A., Wagner, H., et al. (2013). The role of C_4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytol.* 197, 177–185. doi: 10.1111/j.1469-8137.2012.04375.x

Harada, H., and Matsuda, Y. (2005). Identification and characterization of a new carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. *Can. J. Bot.* 83, 909–916. doi: 10.1139/b05-078

Harada, H., Nakatsuma, D., Ishida, M., and Matsuda, Y. (2005). Regulation of the expression of intracellular beta-carbonic anhydrase in response to CO_2 and light in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 139, 1041–1050. doi: 10.1104/pp.105.065185

Harada, H., Nakajima, K., Sakaue, K., and Matsuda, Y. (2006). CO_2 sensing at ocean surface mediated by cAMP in a marine diatom. *Plant Physiol.* 142, 1318–1328. doi: 10.1104/pp.106.086561

Harrison, P. J., Thompson, P. A., and Calderwood, G. S. (1990). Effects of nutrient and light limitation on the biochemical composition of phytoplankton. *J. Appl. Phycol.* 2, 45–56. doi: 10.1007/BF02179768

Hauser, M., Eichelmann, H., Oja, V., Heber, U., and Laisk, A. (1995). Stimulation by light of rapid pH regulation in the chloroplast stroma *in vivo* as indicated by CO_2 solubilization in leaves. *Plant Physiol.* 108, 1059. doi: 10.1104/pp.108.3.1059

Hennon, G. M. M., Ashworth, J., Groussman, R. D., Berthiaume, C., Morales, R. L., Baliga, N. S., et al. (2015). Diatom acclimation to elevated CO_2 via cAMP signalling and coordinated gene expression. *Nat. Clim. Change* 5, 761–765. doi: 10.1038/nclimate2683

Herve, V., Derr, J., Douady, S., Quinet, M., Moisan, L., and Lopez, P. J. (2012). Multiparametric analyses reveal the pH-dependence of silicon biomineralization in diatoms. *PloS One* 7, e46722. doi: 10.1371/journal.pone.0046722

Hesse, S. J. A., Ruijter, G. J. G., Dijkema, C., and Visser, J. (2000). Measurement of intracellular (compartmental) pH by ^{31}P NMR in. *Aspergillus Niger J. Biotechnol.* 77 , 5–15. doi: 10.1016/S0168-1656(99)00203-5

Heydarizadeh, P., Boureba, W., Zahedi, M., Huang, B., Moreau, B., Lukomska, E., et al. (2017). Response of CO_2-starved diatom *Phaeodactylum tricornutum* to light intensity transition. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160396. doi: 10.1098/rstb.2016.0396

Heydarizadeh, P., Veidl, B., Huang, B., Lukomska, E., Wielgosz-Collin, G., Couzinet-Mossion, A., et al. (2019). Carbon orientation in the diatom *Phaeodactylum tricornutum*: the effects of carbon limitation and photon flux density. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00471

Hildebrand, M., Manandhar-Shrestha, K., and Abbriano, R. (2017). Effects of chrysolaminarin synthase knockdown in the diatom *Thalassiosira pseudonana*: Implications of reduced carbohydrate storage relative to green algae. *Algal Res.* 23, 66–77. doi: 10.1016/j.algal.2017.01.010

Hildebrand, M., Lerch, S. J. L., and Shrestha, R. P. (2018). Understanding diatom cell wall silicification—moving forward. *Front. Mar. Sci.* 5. doi: 10.3389/fmars.2018.00125

Hitchcock, G. L. (1980). Diel variation in chlorophyll *a*, carbohydrate and protein content of the marine diatom *Skeletonema costatum*. *Mar. Biol.* 57, 271–278. doi: 10.1007/BF00387570

Hopes, A., Nekrasov, V., Kamoun, S., and Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods* 12, 49. doi: 10.1186/s13007-016-0148-0

Hopkinson, B. M., Dupont, C. L., Allen, A. E., and Morel, F. M. M. (2011). Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3830–3837. doi: 10.1073/pnas.1018062108

Hopkinson, B. M., Dupont, C. L., and Matsuda, Y. (2016). The physiology and genetics of CO_2 concentrating mechanisms in model diatoms. *Curr. Opin. Plant Biol.* 31, 51–57. doi: 10.1016/j.pbi.2016.03.013

Huang, W., and Daboussi, F. (2017). Genetic and metabolic engineering in diatoms. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160411. doi: 10.1098/rstb.2016.0411

Huang, W., Haferkamp, I., Lepetit, B., Molchanova, M., Hou, S., Jeblick, W., et al. (2018). Reduced vacuolar β-1,3-glucan synthesis affects carbohydrate metabolism as well as plastid homeostasis and structure in *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4791–4796. doi: 10.1073/pnas.1719274115

Huysman, M. J. J., Fortunato, A. E., Matthijs, M., Costa, B. S., Vanderhaeghen, R., Van Den Daele, H., et al. (2013). AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *Plant Cell* 25, 215–228. doi: 10.1105/tpc.112.106377

IPCC (2014). "Climate Change. Synthesis Report," in *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. 151. Eds. Core Writing Team, R. K. Pachauri, L. A. Meyer (Geneva, Switzerland: IPCC).

Jensen, E., Clement, R., Maberly, S. C., and Gontero, B. (2017). Regulation of the Calvin-Benson-Bassham cycle in the enigmatic diatoms: biochemical and evolutionary variations on an original theme. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160401. doi: 10.1098/rstb.2016.0401

Jensen, E. L., Clement, R., Kosta, A., Maberly, S. C., and Gontero, B. (2019a). A new widespread subclass of carbonic anhydrase in marine phytoplankton. *ISME J.* 13, 2094–2106. doi: 10.1038/s41396-019-0426-8

Jensen, E. L., Yanguez, K., Carriere, F., and Gontero, B. (2019b). Storage compound accumulation in diatoms as response to elevated CO_2 concentration. *Biology* 9, 5. doi: 10.3390/biology9010005

Jensen, E. L., Maberly, S. C., and Gontero, B. (2020). Insights on the functions and ecophysiological relevance of the diverse carbonic anhydrases in microalgae. *Int. J. Mol. Sci.* 21, 2922. doi: 10.3390/ijms21082922

Keeling, P. J. (2009). Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J. Eukaryot. Microbiol.* 56, 1–8. doi: 10.1111/j.1550-7408.2008.00371.x

Kikutani, S., Tanaka, R., Yamazaki, Y., Hara, S., Hisabori, T., Kroth, P. G., et al. (2012). Redox regulation of carbonic anhydrases via thioredoxin in chloroplast of the marine diatom *Phaeodactylum tricornutum*. *J. Biol. Chem.* 287, 20689–20700. doi: 10.1074/jbc.M111.322743

Kim, J.-M., Lee, K., Shin, K., Kang, J.-H., Lee, H.-W., Kim, M., et al. (2006). The effect of seawater CO_2 concentration on growth of a natural phytoplankton assemblage in a controlled mesocosm experiment. *Limnol. Oceanogr.* 51, 1629–1636. doi: 10.4319/lo.2006.51.4.1629

Knuesting, J., and Scheibe, R. (2018). Small molecules govern thiol redox switches. *Trends Plant Sci.* 23, 769–782. doi: 10.1016/j.tplants.2018.06.007

Kojadinovic-Sirinelli, M., Villain, A., Puppo, C., Sing, S. F., Prioretti, L., Hubert, P., et al. (2018). Exploring the microbiome of the "star" freshwater diatom *Asterionella formosa* in a laboratory context. *Environ. Microbiol.* 20, 3601–3615. doi: 10.1111/1462-2920.14337

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 3, e1426. doi: 10.1371/journal.pone.0001426

Kroth, P. G., Bones, A. M., Daboussi, F., Ferrante, M. I., Jaubert, M., Kolot, M., et al. (2018). Genome editing in diatoms: achievements and goals. *Plant Cell. Rep.* 37, 1401–1408. doi: 10.1007/s00299-018-2334-1

Kuczynska, P., Jemiola-Rzeminska, M., and Strzalka, K. (2015). Photosynthetic pigments in diatoms. *Mar. Drugs* 13, 5847–5881. doi: 10.3390/md13095847

Kustka, A. B., Milligan, A. J., Zheng, H., New, A. M., Gates, C., Bidle, K. D., et al. (2014). Low CO_2 results in a rearrangement of carbon metabolism to support

C_4 photosynthetic carbon assimilation in *Thalassiosira pseudonana*. *New Phytol.* 204, 507–520. doi: 10.1111/nph.12926

Lauritano, C., Ferrante, M. I., and Rogato, A. (2019). Marine natural products from microalgae: an -omics overview. *Mar. Drugs* 17, 269. doi: 10.3390/md17050269

Lavaud, J., Materna, A. C., Sturm, S., Vugrinec, S., and Kroth, P. G. (2012). Silencing of the violaxanthin de-epoxidase gene in the diatom *Phaeodactylum tricornutum* reduces diatoxanthin synthesis and non-photochemical quenching. *PloS One* 7, e36806. doi: 10.1371/journal.pone.0036806

Lin, Q., Liang, J. R., Huang, Q. Q., Luo, C. S., Anderson, D. M., Bowler, C., et al. (2017). Differential cellular responses associated with oxidative stress and cell fate decision under nitrate and phosphate limitations in *Thalassiosira pseudonana*: Comparative proteomics. *PloS One* 12, e0184849. doi: 10.1371/journal.pone.0184849

Lohr, M., and Wilhelm, C. (1999). Algae displaying the diadinoxanthin cycle also possess the violaxanthin cycle. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8784–8789. doi: 10.1073/pnas.96.15.8784

Loiselle, F. B., and Casey, J. R. (2010). "Measurement of intracellular pH," in *Membrane transporters in drug discovery and development: methods and protocols*. Ed. Q. Yan (Totowa, NJ: Humana Press), 311–331.

Maberly, S. C., and Gontero, B. (2017). Ecological imperatives for aquatic CO_2-concentrating mechanisms. *J. Exp. Bot.* 68, 3797–3814. doi: 10.1093/jxb/erx201

Maberly, S. C., Courcelle, C., Groben, R., and Gontero, B. (2010). Phylogenetically-based variation in the regulation of the Calvin cycle enzymes, phosphoribulokinase and glyceraldehyde-3-phosphate dehydrogenase, in algae. *J. Exp. Bot.* 61, 735–745. doi: 10.1093/jxb/erp337

Maier, T., Güell, M., and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* 583, 3966–3973. doi: 10.1016/j.febslet.2009.10.036

Mann, D. G., and Vanormelingen, P. (2013). An inordinate fondness? The number, distributions, and origins of diatom species. *J. Eukaryot. Microbiol.* 60, 414–420. doi: 10.1111/jeu.12047

Marri, L., Thieulin-Pardo, G., Lebrun, R., Puppo, R., Zaffagnini, M., Trost, P., et al. (2014). CP12-mediated protection of Calvin-Benson cycle enzymes from oxidative stress. *Biochimie* 97, 228–237. doi: 10.1016/j.biochi.2013.10.018

Matsuda, Y., and Kroth, P. G. (2014). ""Carbon fixation in diatoms,"," in *The structural basis of biological energy generation. Advances in photosynthesis and respiration (including bioenergy and related processes)*. Ed. M. F. Hohmann-Marriott (Dordrecht: Springer Netherlands), 335–362.

Matsuda, Y., Nakajima, K., and Tachibana, M. (2011). Recent progresses on the genetic basis of the regulation of CO_2 acquisition systems in response to CO_2 concentration. *Photosynth. Res.* 109, 191–203. doi: 10.1007/s11120-011-9623-7

Matsuda, Y., Hopkinson, B. M., Nakajima, K., Dupont, C. L., and Tsuji, Y. (2017). Mechanisms of carbon dioxide acquisition and CO_2 sensing in marine diatoms: a gateway to carbon metabolism. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160403. doi: 10.1098/rstb.2016.0403

Medlin, L. K. (2016). Evolution of the diatoms: major steps in their evolution and a review of the supporting molecular and morphological evidence. *Phycologia* 55, 79–103. doi: 10.2216/15-105.1

Mekhalfi, M., Avilan, L., Lebrun, R., Botebol, H., and Gontero, B. (2012). Consequences of the presence of 24-epibrassinolide, on cultures of a diatom, *Asterionella formosa*. *Biochimie* 94, 1213–1220. doi: 10.1016/j.biochi.2012.02.011

Mekhalfi, M., Puppo, C., Avilan, L., Lebrun, R., Mansuelle, P., Maberly, S. C., et al. (2014). Glyceraldehyde-3-phosphate dehydrogenase is regulated by ferredoxin-NADP reductase in the diatom *Asterionella formosa*. *New Phytol.* 203, 414–423. doi: 10.1111/nph.12820

Michels, A. K., Wedel, N., and Kroth, P. G. (2005). Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Physiol.* 137, 911–920. doi: 10.1104/pp.104.055285

Mock, T., Otillar, R. P., Strauss, J., Mcmullan, M., Paajanen, P., Schmutz, J., et al. (2017). Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. *Nature* 541, 536–540. doi: 10.1038/nature20803

Montrichard, F., Alkhalfioui, F., Yano, H., Vensel, W. H., Hurkman, W. J., and Buchanan, B. B. (2009). Thioredoxin targets in plants: the first 30 years. *J. Proteomics* 72, 452–474. doi: 10.1016/j.jprot.2008.12.002

Morel, F. M. M., Lam, P. J., and Saito, M. A. (2020). Trace metal substitution in marine phytoplankton. *Annu. Rev. Earth Planet. Sci.* 48, 491–517. doi: 10.1146/annurev-earth-053018-060108

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324, 1724–1726. doi: 10.1126/science.1172983

Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA^+ protein CbbX, a red-type Rubisco activase. *Nature* 479, 194–199. doi: 10.1038/nature10568

Muhseen, Z. T., Xiong, Q., Chen, Z., and Ge, F. (2015). Proteomics studies on stress responses in diatoms. *Proteomics* 15, 3943–3953. doi: 10.1002/pmic.201500165

Nakajima, K., Tanaka, A., and Matsuda, Y. (2013). SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1767–1772. doi: 10.1073/pnas.1216234110

Nikolaev, V. O., and Lohse, M. J. (2006). Monitoring of cAMP synthesis and degradation in living cells. *Physiology* 21, 86–92. doi: 10.1152/physiol.00057.2005

Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., et al. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules* 9, 322. doi: 10.3390/biom9080322

Norici, A., Bazzoni, A. M., Pugnetti, A., Raven, J. A., and Giordano, M. (2011). Impact of irradiance on the C allocation in the coastal marine diatom *Skeletonema marinoi* Sarno and Zingone. *Plant Cell Environ.* 34, 1666–1677. doi: 10.1111/j.1365-3040.2011.02362.x

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6, 24951. doi: 10.1038/srep24951

Ogura, A., Akizuki, Y., Imoda, H., Mineta, K., Gojobori, T., and Nagai, S. (2018). Comparative genome and transcriptome analysis of diatom, *Skeletonema costatum*, reveals evolution of genes for harmful algal bloom. *BMC Genomics* 19, 765. doi: 10.1186/s12864-018-5144-5

Ohno, N., Inoue, T., Yamashiki, R., Nakajima, K., Kitahara, Y., Ishibashi, M., et al. (2012). CO_2-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-Like basic zipper domain in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 158, 499–513. doi: 10.1104/pp.111.190249

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. *Mol. Genet. Genomics* 277, 427–439. doi: 10.1007/s00438-006-0199-4

Perez-Perez, M. E., Mauries, A., Maes, A., Tourasse, N. J., Hamon, M., Lemaire, S. D., et al. (2017). The deep thioredoxome in *Chlamydomonas reinhardtii*: new insights into redox regulation. *Mol. Plant* 10, 1107–1125. doi: 10.1016/j.molp.2017.07.009

Ponnala, L., Wang, Y., Sun, Q., and Van Wijk, K. J. (2014). Correlation of mRNA and protein abundance in the developing maize leaf. *Plant J.* 78, 424–440. doi: 10.1111/tpj.12482

Portis, A. R., and Heldt, H. W. (1976). Light-dependent changes of Mg^{2+} concentration in stroma in relation to Mg^{2+} dependency of CO_2 fixation in intact chloroplasts. *BBA-Bioenergetics* 449, 434–446. doi: 10.1016/0005-2728(76)90154-7

Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wires Syst. Biol. Med.* 4, 565–583. doi: 10.1002/wsbm.1185

Prosser, G. A., Larrouy-Maumus, G., and De Carvalho, L. P. (2014). Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO Rep.* 15, 657–669. doi: 10.15252/embr.201338283

Pupillo, P., and Giulianipiccari, G. (1975). The reversible depolymerization of spinach chloroplast glyceraldehyde-phosphate dehydrogenase-interaction with nucleotides and dithiothreitol. *Eur. J. Biochem.* 51, 475–482. doi: 10.1111/j.1432-1033.1975.tb03947.x

Raven, J., Brown, K., Mackay, M., Beardall, J., Giordano, M., Granum, E., et al. (2005). "Iron, nitrogen, phosphorus and zinc cycling and consequences for primary productivity in the oceans," in *Micro-organisms and earth systems-*

advances in geomicrobiology. Eds. G. Gadd, K. Semple and H. Lappin-Scott (Cambridge, UK: Cambridge University Press), 247–272.

Raven, J. A. (1994). Carbon fixation and carbon availability in marine phytoplankton. *Photosynth. Res.* 39, 259–273. doi: 10.1007/BF00014587

Reinfelder, J. R., Kraepiel, A. M. L., and Morel, F. M. M. (2000). Unicellular C_4 photosynthesis in a marine diatom. *Nature* 407, 996–999. doi: 10.1038/35039612

Reinfelder, J. R. (2011). "Carbon concentrating mechanisms in eukaryotic marine phytoplankton," in *Annu. Rev. Mar. Sci*. Eds. C. A. Carlson and S. J. Giovannoni (Palo Alto, USA: Annual Reviews Publisher), 291–315.

Riebesell, U., Wolfgladrow, D. A., and Smetacek, V. (1993). Carbon dioxide limitation of marine phytoplankton growth rates. *Nature* 361, 249–251. doi: 10.1038/361249a0

Rosenwasser, S., Graff Van Creveld, S., Schatz, D., Malitsky, S., Tzfadia, O., Aharoni, A., et al. (2014). Mapping the diatom redox-sensitive proteome provides insight into response to nitrogen stress in the marine environment. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2740–2745. doi: 10.1073/pnas.1319773111

Sage, R. F. (2004). The evolution of C_4 photosynthesis. *New Phytol.* 161, 341–370. doi: 10.1111/j.1469-8137.2004.00974.x

Samukawa, M., Shen, C., Hopkinson, B. M., and Matsuda, Y. (2014). Localization of putative carbonic anhydrases in the marine diatom, Thalassiosira pseudonana. *Photosynth. Res.* 121, 235–249. doi: 10.1007/s11120-014-9967-x

Santin, Y. G., Doan, T., Lebrun, R., Espinosa, L., Journet, L., and Cascales, E. (2018). In vivo TssA proximity labelling during type VI secretion biogenesis reveals TagA as a protein that stops and holds the sheath. *Nat. Microbiol.* 3, 1304–1313. doi: 10.1038/s41564-018-0234-3

Satoh, D., Hiraoka, Y., Colman, B., and Matsuda, Y. (2001). Physiological and molecular biological characterization of intracellular carbonic anhydrase from the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 126, 1459–1470. doi: 10.1104/pp.126.4.1459

Schoefs, B., Hu, H., and Kroth, P. G. (2017). The peculiar carbon metabolism in diatoms. *Philos. Trans. R. Soc. B Biol. Sci.* 372, 20160405. doi: 10.1098/rstb.2016.0405

Schurmann, P., and Jacquot, J. P. (2000). Plant thioredoxin systems revisited. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 371–400. doi: 10.1146/annurev.arplant.51.1.371

Stein, K. C., and Frydman, J. (2019). The stop-and-go traffic regulating protein biogenesis: how translation kinetics controls proteostasis. *J. Biol. Chem.* 294, 2076–2084. doi: 10.1074/jbc.REV118.002814

Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., and Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* 5, 5764. doi: 10.1038/ncomms6764

Subramanian, S., Barry, A. N., Pieris, S., and Sayre, R. T. (2013). Comparative energetics and kinetics of autotrophic lipid and starch metabolism in chlorophytic microalgae: implications for biomass and biofuel production. *Biotechnol. Biofuels* 6, 150. doi: 10.1186/1754-6834-6-150

Sun, N., Ma, L. G., Pan, D. Y., Zhao, H. Y., and Deng, X. W. (2003). Evaluation of light regulatory potential of Calvin cycle steps based on large-scale gene expression profiling data. *Plant Mol. Biol.* 53, 467–478. doi: 10.1023/B:PLAN.0000019071.12878.9e

Tachibana, M., Allen, A. E., Kikutani, S., Endo, Y., Bowler, C., and Matsuda, Y. (2011). Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynth. Res.* 109, 205–221. doi: 10.1007/s11120-011-9634-4

Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Marechal, E., et al. (2015). Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* 27, 162–176. doi: 10.1105/tpc.114.135194

Tanaka, A., Ohno, N., Nakajima, K., and Matsuda, Y. (2016). Light and CO_2/cAMP signal cross talk on the promoter elements of chloroplastic beta-carbonic anhydrase genes in the marine diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 170, 1105–1116. doi: 10.1104/pp.15.01738

Taraldsvik, M., and Myklestad, S. M. (2000). The effect of pH on growth rate, biochemical composition and extracellular carbohydrate production of the marine diatom *Skeletonema costatum*. *Eur. J. Phycol.* 35, 189–194. doi: 10.1080/09670260010001735781

Thieulin-Pardo, G., Remy, T., Lignon, S., Lebrun, R., and Gontero, B. (2015). Phosphoribulokinase from *Chlamydomonas reinhardtii*: a Benson-Calvin cycle enzyme enslaved to its cysteine residues. *Mol. Biosyst.* 11, 1134–1145. doi: 10.1039/C5MB00035A

Torstensson, A., Chierici, M., and Wulff, A. (2012). The influence of increased temperature and carbon dioxide levels on the benthic/sea ice diatom *Navicula directa*. *Polar Biol.* 35, 205–214. doi: 10.1007/s00300-011-1056-4

Tortell, P. D. (2000). Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol. Oceanogr.* 45, 744–750. doi: 10.4319/lo.2000.45.3.0744

Traller, J. C., Cokus, S. J., Lopez, D. A., Gaidarenko, O., Smith, S. R., Mccrow, J. P., et al. (2016). Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol. Biofuels* 9, 258. doi: 10.1186/s13068-016-0670-3

Tsuji, Y., Nakajima, K., and Matsuda, Y. (2017). Molecular aspects of the biophysical CO_2-concentrating mechanism and its regulation in marine diatoms. *J. Exp. Bot.* 68, 3763–3772. doi: 10.1093/jxb/erx173

Valegard, K., Andralojc, P. J., Haslam, R. P., Pearce, F. G., Eriksen, G. K., Madgwick, P. J., et al. (2018). Structural and functional analyses of Rubisco from arctic diatom species reveal unusual posttranslational modifications. *J. Biol. Chem.* 293, 13033–13043. doi: 10.1074/jbc.RA118.003518

Valenzuela, J. J., Garcia De Lomana, A. L., Lee, A., Armbrust, E. V., Orellana, M. V., and Baliga, N. S. (2018). Ocean acidification conditions increase resilience of marine diatoms. *Nat. Commun.* 9, 2328. doi: 10.1038/s41467-018-04742-3

Van Der Star, W. R., Dijkema, C., De Waard, P., Picioreanu, C., Strous, M., and Van Loosdrecht, M. C. (2010). An intracellular pH gradient in the anammox bacterium *Kuenenia stuttgartiensis* as evaluated by ^{31}P NMR. *Appl. Microbiol. Biot.* 86, 311–317. doi: 10.1007/s00253-009-2309-9

Villain, A., Kojadinovic, M., Puppo, C., Prioretti, L., Hubert, P., Zhang, Y., et al. (2017). Complete mitochondrial genome sequence of the freshwater diatom *Asterionella formosa*. *Mitochondrial DNA B* 2, 97–98. doi: 10.1080/23802359.2017.1285210

Weber, T., Gruber, A., and Kroth, P. G. (2009). The presence and localization of thioredoxins in diatoms, unicellular algae of secondary endosymbiotic origin. *Mol. Plant* 2, 468–477. doi: 10.1093/mp/ssp010

Wei, L., El Hajjami, M., Shen, C., You, W., Lu, Y., Li, J., et al. (2019). Transcriptomic and proteomic responses to very low CO_2 suggest multiple carbon concentrating mechanisms in *Nannochloropsis oceanica*. *Biotechnol. Biofuels* 12, 168. doi: 10.1186/s13068-019-1506-8

Werdan, K., and Heldt, H. W. (1973). Measurement of pH value in stroma and thylakoid space of intact chloroplasts. *Hoppe Seylers Z. Physiol. Chem.* 354, 223–224.

Werdan, K., Heldt, H. W., and Milovancev, M. (1975). The role of pH in regulation of carbon fixation in chloroplast stroma. Studies on CO_2 fixation in light and dark. *BBA-Bioenergetics* 396, 276–292. doi: 10.1016/0005-2728(75)90041-9

Wilhelm, C., Buechel, C., Fisahn, J., Goss, R., Jakob, T., Laroche, J., et al. (2006). The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157, 91–124. doi: 10.1016/j.protis.2006.02.003

Young, J. N., and Morel, F. M. M. (2015). The CO_2 switch in diatoms. *Nat. Clim. Change* 5, 722–723. doi: 10.1038/nclimate2691

Young, J. N., Heureux, A. M. C., Sharwood, R. E., Rickaby, R. E. M., Morel, F. M. M., and Whitney, S. M. (2016). Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *J. Exp. Bot.* 67, 3445–3456. doi: 10.1093/jxb/erw163

Zaffagnini, M., Bedhomme, M., Groni, H., Marchand, C. H., Puppo, C., Gontero, B., et al. (2012). Glutathionylation in the photosynthetic model organism *Chlamydomonas reinhardtii*: a proteomic survey. *Mol. Cell. Proteomics* 11, M111.014142. doi: 10.1074/mcp.M111.014142

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Launay, Huang, Maberly and Gontero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 16 October 2020
doi: 10.3389/fpls.2020.590949

PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom *Phaeodactylum tricornutum*

Ouardia Ait-Mohamed[1†], Anna M. G. Novák Vanclová[1‡], Nathalie Joli[1‡], Yue Liang[2], Xue Zhao[1,3], Auguste Genovesio[1], Leila Tirichine[1,3], Chris Bowler[1]* and Richard G. Dorrell[1]*

[1] Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France, [2] Department of Oceanography, Dalhousie University, Halifax, NS, Canada, [3] Université de Nantes, CNRS, UFIP, UMR 6286, Nantes, France

Transcriptional coordination is a fundamental component of prokaryotic and eukaryotic cell biology, underpinning the cell cycle, physiological transitions, and facilitating holistic responses to environmental stress, but its overall dynamics in eukaryotic algae remain poorly understood. Better understanding of transcriptional partitioning may provide key insights into the primary metabolism pathways of eukaryotic algae, which frequently depend on intricate metabolic associations between the chloroplasts and mitochondria that are not found in plants. Here, we exploit 187 publically available RNAseq datasets generated under varying nitrogen, iron and phosphate growth conditions to understand the co-regulatory principles underpinning transcription in the model diatom *Phaeodactylum tricornutum*. Using WGCNA (Weighted Gene Correlation Network Analysis), we identify 28 merged modules of co-expressed genes in the *P. tricornutum* genome, which show high connectivity and correlate well with previous microarray-based surveys of gene co-regulation in this species. We use combined functional, subcellular localization and evolutionary annotations to reveal the fundamental principles underpinning the transcriptional co-regulation of genes implicated in *P. tricornutum* chloroplast and mitochondrial metabolism, as well as the functions of diverse transcription factors underpinning this co-regulation. The resource is publically available as PhaeoNet, an advanced tool to understand diatom gene co-regulation.

Keywords: stramenopile, transcriptomics, sigma factors, aureochromes, epigenetics, chloroplast-mitochondria

OPEN ACCESS

Edited by:
Justine Marchand, Le Mans Université, France

Reviewed by:
Pannaga Pavan Jutur, International Centre for Genetic Engineering and Biotechnology, India
Yoshihisa Hirakawa, University of Tsukuba, Japan

***Correspondence:**
Leila Tirichine
Leila.Tirichine@univ-nantes.fr
Chris Bowler
cbowler@biologie.ens.fr

[†]***Present address:***
Ouardia Ait-Mohamed, Immunity and Cancer Department, Institut Curie, PSL Research University, INSERM U932, Paris, France

[‡]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *03 August 2020*
Accepted: *15 September 2020*
Published: *16 October 2020*

Citation:
Ait-Mohamed O, Novák Vanclová AMG, Joli N, Liang Y, Zhao X, Genovesio A, Tirichine L, Bowler C and Dorrell RG (2020) PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom Phaeodactylum tricornutum. Front. Plant Sci. 11:590949. doi: 10.3389/fpls.2020.590949

INTRODUCTION

The biology of prokaryotic and eukaryotic cells is dependent on elaborate metabolic, regulatory and gene expression pathways, consisting of multiple interacting components. The successful operation of these pathways depend on the coordinated expression of genes that underpin them, which allow the stoichiometric assembly of their constituent components and enable discrete and

appropriate regulatory responses to changes in physiological conditions (Gasch and Eisen, 2002; Teichmann and Babu, 2002). In prokaryotes and bacteria-derived genomes (e.g., "plastids" including chloroplasts and mitochondria) gene co-regulation is possible via the co-transcription of linked genes as part of the same transcriptional operon (Teichmann and Babu, 2002). In contrast, gene order plays a limited role in eukaryotic nuclear gene co-expression (Michalak, 2008), which depends instead on the simultaneous transcription, or transcriptional stabilization, of multiple discrete genomic loci. This may occur through common transcription factors (Teichmann and Babu, 2002; Reja et al., 2015); epigenetic modifications based around characteristic histone and DNA marks (Bird, 2002; Bártová et al., 2008); co-ordinated transcript processing events (Norbury, 2010); and the activity of small and long regulatory non-coding RNAs (Tsai et al., 2010; Kim and Sung, 2012).

The degree to which gene co-regulation is shared between different species is debated, with different studies identifying shared co-regulatory trends in between 8% (Teichmann and Babu, 2002) and 70% (Snel et al., 2004) of orthologous gene pairs between *Saccharomyces cerevisiae* (yeast) and *Caenorhabditis elegans* (nematode) genomes. Nonetheless, there is substantial merit to understanding gene co-regulation patterns in novel species. Since their origins over two billion years ago, the eukaryotes have radiated into a diverse range of different lineages, many of which are unicellular; and distantly related to model organisms in the animals, fungi and plants, with different underlying cellular biology and transcriptional dynamics (Walker et al., 2011). Identifying what gene co-regulation processes occur in microbial eukaryotes may allow us to better understand the biology underpinning the base of planetary food webs; and to better model the robustness of eukaryotic communities to environmental change.

The diatoms are a major group of predominantly marine algae, believed to be responsible for nearly one-fifth of total planetary photosynthesis (Field et al., 1998). Diatoms sit within the stramenopile supergroup, and are distantly related to animals, fungi and plants. Photosynthetic members of the stramenopiles, including diatoms, possess a chloroplast acquired through secondary endosymbiosis, unlike the primary plant chloroplast, which is of primary endosymbiotic origin (Walker et al., 2011). Previous genomic and functional investigations of model diatoms, for example *Phaeodactylum tricornutum*, have identified divergent features in their cellular biology, compared to more well-understood eukaryotic groups (Bowler et al., 2010). These include intricate metabolic connections between the diatom chloroplast, mitochondria and cytoplasm (Prihoda et al., 2012; Bailleul et al., 2015); and a wide range of different histone structural modifications (Veluchamy et al., 2013, 2015), many of which have not yet been detected in more established eukaryotic models.

Previously, microarray data from over 100 different conditions, including illumination regimes and pollutant stress (Osborn and Hook, 2013; Valle et al., 2014), have been generated from *P. tricornutum*; which have been assembled into a searchable interface, DiatomPortal that divides the *P. tricornutum* genome into 500 co-regulated gene clusters (Ashworth et al., 2016). Alongside this, a suite of RNA sequencing libraries exploring cellular responses to phosphorus, iron and nitrogen limitation have now been generated (Maheswari et al., 2010; Cruz de Carvalho et al., 2016; Smith et al., 2016; McCarthy et al., 2017) and inspecting these data may allow more precise integration of quantitative differences in transcript abundance than would be possible through microarray analyses. Furthermore, co-expression networks are a powerful tool for functional prediction and annotation of unknown genes in the absence of prior knowledge, which is the case for a significant number of genes in *P. tricornutum* (Rastogi et al., 2018). Co-expression networks can furthermore enrich our understanding of the more sparse co-expression networks generated for other marine algal species with secondary chloroplasts (principally, the distantly related diatom *Thalassiosira pseudonana*, the distantly related stramenopile *Nannochloropsis oceanica* and the haptophyte *Emiliania huxleyi*; Ashworth et al., 2016; Ashworth and Ralph, 2018).

Here, we use a tool of gene co-expression network analysis, WGCNA (Weighted Gene Correlation Network Analysis (Langfelder and Horvath, 2008; Guidi et al., 2016), to build PhaeoNet, an advanced tool for transcriptional understanding of the *P. tricornutum* genome. PhaeoNet is composed of 28 co-regulated gene modules, each with different expression dynamics. Considering the repartition of genes within these modules; functional, epigenetic and localization information from the third version annotation of the *P. tricornutum* genome (Phatr3; Rastogi et al., 2018); and annotated lists of diatom transcription factors (Rayko et al., 2010), we identify core features underpinning the transcriptional partitioning of diatom primary metabolism, including probable metabolic links between the diatom mitochondria and chloroplast; and dissect the diverse ranges of different transcriptional drivers of this co-regulation, notably in the case of chloroplast-targeted sigma factors. The raw data underpinning PhaeoNet have been made publically accessible via https://osf.io/42xmp.

MATERIALS AND METHODS

Dataset Curation and Abundance Calculations

A total of 187 publically available RNA-seq datasets from *P. tricornutum*, generated from three studies exploring, respectively, phosphorus (Cruz de Carvalho et al., 2016), iron (Smith et al., 2016) and nitrogen (McCarthy et al., 2017) stress transcriptional responses, were collected from the sequence read archive (SRA)[1] (Wheeler et al., 2006). The 182 libraries that passed through quality control steps, were included in the final version of the WGCNA performed, are named per their names respective studies in **Supplementary Table 1**, sheet 1. Data provided in the phosphate and nitrogen conditions were obtained using an Illumina Genome Analyzer (Bentley et al., 2008), while the iron study used SOLiD technology sequencing (Morey et al., 2013). *P. tricornutum* transcript IDs from each study were mapped to gene models based on the

[1]http://www.ncbi.nlm.nih.gov/Traces/sra/

Phatr3 annotation of the genome (**Supplementary Table 1**, Sheet 2; Rastogi et al., 2018).

Raw data were reprocessed using FastQC version v0.11.5[2]. Low quality reads (Phred quality score below 20) were filtered-out using trim-galore version 0.5.0[3]. The remaining sequences were aligned to the reference genome with the software package STAR version 2.5.3a (Dobin et al., 2013) (STAR –outFilterMismatchNmax 2 –outFilterMultimapNmax 1000 –alignIntronMin 20 –alignIntronMax 2000). The iron data derived from the SOLiD technology were first mapped using the Life Technologies LifeScope software suitable for data from such technology. For homogeneity purposes, the reads were remapped using the pre-cited version of STAR.

Expression levels of individual genes were obtained using featureCounts version 1.6.1 (Liao et al., 2014). Quality checks of datasets were performed using methods provided in DESeq2 version 1.19.37 (Love et al., 2014), with a PCA projection and a hierarchical dendrogram using Spearman correlation between library-normalized gene counts (Glasser and Winter, 1961). These subsequent analyses and results visualizations were performed using R package version 3.4.4.

Weighted Gene Correlation Network Analysis (WGCNA) and Network Visualization

The WGCNA R package (Langfelder and Horvath, 2008) was used to identify network modules from library-normalized gene expression values. First, a signed adjacency matrix (accepting oppositely correlated gene expression values to be clustered in the same modules) was obtained by calculating the pairwise Bi-weight mid-correlation coefficient from rij (Langfelder and Horvath, 2008), that represent expression values of genes i and j. A connectivity measure (k) per gene set was calculated by summing the connection strengths with other gene sets. Subsequently, the weighted adjacency matrix was obtained by raising the absolute value of the pairwise gene expression correlations to the soft-thresholding parameter β (Zhao et al., 2010). This achieved the scale-free topology criterion for WGCNA and typical for biological networks, emphasizing high correlations and minoring low ones, in which most nodes are not connected and only a few nodes are highly connected (Barabási, 2009).

The scale-free topology of PhaeoNet was evaluated by the Scale-Free Topology Fitting Index (R^2), which was the square of the correlation between log[p(k)] and log(k). A β coefficient of 12 with R^2 of 0.9 was used during the network building from the signed weighted adjacency matrix. The weighted adjacency matrix was finally used to calculate the Topological Overlap Matrix (TOM). Subsequently, modules were detected on the basis of the Topological Overlap measure using the following parameters: minModuleSize = 40 and mergeCutHeight = 0.25.

Graphical representations of the network were performed using Cytoscape (Shannon et al., 2003). All code used for the construction of PhaeoNet and interactive diagrams of each merged module are publically available through the following link: https://osf.io/42xmp.

Biological Interpretation of Merged Modules

The distribution of *P. tricornutum* genes in each transcriptional module was compared to the distribution of orthologous gene models (Phatr2.0 genome annotation) in microarray-derived transcriptional clusters generated as part of the DiatomPortal project (Ashworth et al., 2016). Only gene models that showed a one-to-one gene mapping (i.e., gene models that were neither split or merged, but including gene models that were truncated or extended) between version 2 (Phatr2) and version 3 (Phatr3) annotations of the *P. tricornutum* genome (Bowler et al., 2008; Rastogi et al., 2018) were considered.

Biological functions within the merged modules were identified using gene functional annotations from the Phatr3 annotation of the *P. tricornutum* genome (Bowler et al., 2008; Rastogi et al., 2018). These included: GO terms, using the R package TopGO (Aibar et al., 2015); PFAM domains and biological processes (Rastogi et al., 2018); probable evolutionary affinities inferred by BLAST top hit analyses (Rastogi et al., 2018); histone and DNA modifications associated with cells grown in replete media (Veluchamy et al., 2013, 2015); Polycomb group protein marks (Zhao et al., 2020); and KEGG orthology predictions, obtained with BLASTkoala, Kofamkoala and GHOSTkoala servers (Moriya et al., 2007; Kanehisa, 2017; Aramaki et al., 2019; Kanehisa and Sato, 2020). *In silico* targeting predictions were performed for all N-complete protein sequences (i.e., protein sequences inferred to start in a methionine) within the dataset, using HECTAR (Gschloessl et al., 2008); ASAFind v2.0 (Gruber et al., 2015), in conjunction with SignalP v3.0 (Bendtsen et al., 2004); MitoFates, with a threshold detection value of 0.35 (Fukasawa et al., 2015; Dorrell et al., 2017); and WolfPSort, taking the consensus best-scoring prediction using animal, fungi and plant reference datasets (Horton et al., 2007). Enrichments in each category were analyzed both qualitatively/manually and by a simple pivot table and chi-squared test. Tabulated lists of all annotations are presented in **Supplementary Table 2**.

Core chloroplast and mitochondria-associated functions were assembled from a list of 524 KEGG ortholog numbers based on previously identified chloroplast and mitochondria functions in photosynthetic eukaryotes (Dorrell et al., 2017; Nonoyama et al., 2019; Novák Vanclová et al., 2020). Where multiple candidate proteins were detected, proteins were assigned to either the chloroplast, mitochondria, or dual chloroplast/mitochondria (Gile et al., 2015; Dorrell et al., 2017) based on *in silico* targeting predictions. Where no clear targeting predictions could be obtained, proteins were identified based on BLAST similarity to orthologous chloroplast- or mitochondria-targeted proteins from other algal and stramenopile species (Dorrell et al., 2017; Río Bártulos et al., 2018). Disregarding 135 query proteins coded by organellar genomes in diatoms (Yu et al., 2018) and 17 query proteins encoded by nuclear genes with no PhaeoNet module

[2]https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[3]http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

assigned, the final set comprised of 372 unique proteins targeted to the chloroplast and/or mitochondrion, encoded by nuclear genes that belong to one of the 28 merged modules. The main metabolic pathways and complexes and quantitative pathway associations, are presented in **Supplementary Table 3**.

A complete list of *P. tricornutum* transcription factors (TF) was assembled from a previous dataset (Rayko et al., 2010) and an updated list specifically of aureochromes (Banerjee et al., 2016), which were mapped to the version 3 genome annotation by BLASTp analysis. A total of 188 candidates, from 18 TF families (HSF, Myb, Zn_finger_C2H2, bZIP, Zn_finger_CCCH, bHLH, Sigma-70, Zn_finger_TAZ, CBF/NF, E2F-DP, CSF, Aureochrome, TRF, CCAAT-binding, AP2-EREBP, TAF9, CXC, Homeobox) corresponded to genes assigned to a PhaeoNet merged module (**Figure 5** and **Supplementary Table 4**). Given that the regulation of gene expression by transcription factors play a key role in the growth and progression of the cell cycle, the distribution within merged modules genes implicated in the cell cycle (cyclins) and in light perception events (e.g., phytochrome, cryptochrome) were additionally investigated, as well as genes implied in transcription and histone-related processes (**Figure 5** and **Supplementary Table 4**; Huysman et al., 2013; Annunziata et al., 2019).

Phylogenetics

A tree of sigma factor proteins from *P. tricornutum* and orthologous diatom and non-diatom sequences was constructed using a pipeline adapted from previous studies (Dorrell et al., 2017; Rastogi et al., 2018). Briefly, the complete peptide sequences of each sigma factor protein (eight total) in the Phatr3 annotation of the *P. tricornutum* genome (Rastogi et al., 2018) were searched using BLASTp against a composite library consisting of 110 diatom genomes, MMETSP (Marine Microbial Eukaryote Transcriptome Sequencing Project, Keeling et al., 2014) and independent transcriptomes; and a reference set of 59 additional eukaryotic genomes, sampled from across the tree of life (**Supplementary Table 5**). Orthologs with an *e*-value of 10^{-05} or lower were extracted and searched against the complete protein sequences encoded within the Phatr3 annotation of the *P. tricornutum* genome via reciprocal BLASTp searches. Sequences which retrieved a single best hit against a *P. tricornutum* sigma factor protein were aligned using MAFFT v 8.0 (Katoh et al., 2002) under the –auto setting (BLOSUM62 matrix, gap open penalty 1.53, offset value 0) and the in-house alignment program in GeneIOUS v 10.0.9 (Kearse et al., 2012) using a more stringent set of conditions (65% similarity cost matrix, gap open penalty 12, gap extension penalty 3, two rounds of refinement). Poorly aligned or incomplete sequences were removed at each step. The 771 protein sequences retained were manually curated to retain a representative series of 86 diatom and non-diatom sequences related to each *P. tricornutum* sigma factor and trimmed using trimal with the –gt 0.5 setting (Capella-Gutiérrez et al., 2009) to yield a 453 aa alignment. The best-scoring tree topology was inferred from the alignment using RAxML v8.2, 100 bootstrap replicates and the PROTGAMMAJTT substitution model (Stamatakis, 2014); and MrBayes v3.2.7 over 600,000 generations, burnin fractions of 0.5 and the Jones amino acid substitution model (Huelsenbeck and Ronquist, 2001). Alignment and tree outputs are provided in **Supplementary Table 5**.

RESULTS AND DISCUSSION

Construction of an Optimized WGCNA Gene Expression Dataset for *P. tricornutum*

We harnessed 187 publically available RNAseq datasets derived from diverse physiological conditions and genotypes (Cruz de Carvalho et al., 2016; Smith et al., 2016; McCarthy et al., 2017) to build an integrative model of gene co-regulation for the model diatom species *P. tricornutum* (**Figure 1A** and **Supplementary Table 1**, sheet 1). We chose to build a dataset focusing on one species only, as even closely related diatom species may contain very different protein orthogroups (Parks et al., 2018; Sato et al., 2020) and even orthologous proteins may perform different physiological functions between different diatom species, with presumably different co-regulatory dynamics (Lampe et al., 2018). *P. tricornutum* was selected as a model system for this study as vastly greater amounts of gene expression data have been generated for this species than any other marine alga (Ashworth and Ralph, 2018); and as its genome annotation (currently in third version form and verified by comparison to over forty RNAseq libraries generated under varied conditions, Rastogi et al., 2018) is arguably the most complete of any alga known, allowing unprecedented insight into protein diversity, including variant protein forms generated by alternative splicing, protein sub-cellular localization and epigenetic modifications. The use of RNAseq data for this analysis allows us to advance on previous (e.g., microarray-based, Ashworth et al., 2016) analyses by allowing us to consider absolute rather than relative changes in expression levels between different datasets, and therefore exclude distorting effects caused by low absolute levels of the expression of specific genes in the *P. tricornutum* genome.

We optimized our data through several key pre-processing steps, for example removing batch effects (**Supplementary Figure 1A**) and five samples showing strong outlier effects (exemplar shown in **Supplementary Figure 1B**) prior to network construction, retaining 182 datasets for the final network construction. We also excluded genes that were found to be lowly expressed (median expression < 10 reads) in all inspected conditions, retaining 10,650/12,177 genes in the Phatr3 annotation (Rastogi et al., 2018) of the *P. tricornutum* genome (**Supplementary Table 1**, sheet 2). All pairwise gene correlations were calculated and then converted into connectivity strengths by raising their values to the power $\beta = 12$ for PhaeoNet. This power makes it possible to work in a scale-free condition and to avoid weak correlations (**Supplementary Figure 2**).

By applying the dynamic tree cut function on the dendrogram obtained by a hierarchical clustering with the method average, we identified 50 WGCNA modules with similar connection force profiles (**Figures 1B,C**). This was reduced to a subset of 28 merged modules with internal correlations above 0.75 (**Figure 1D**, **Supplementary Table 1**,

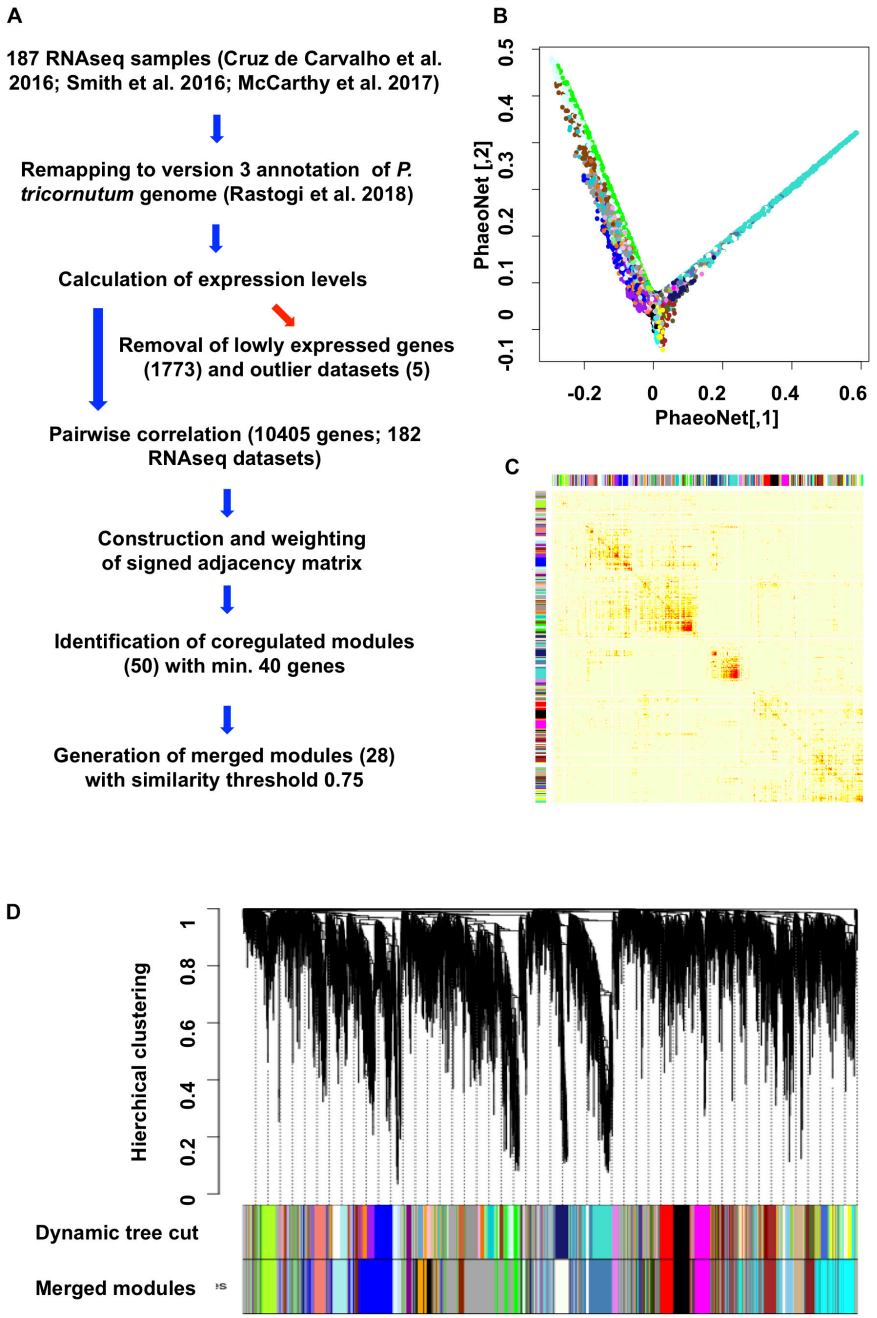


FIGURE 1 | Construction and topology of PhaeoNet. **(A)** Workflow diagram of the steps performed to construct PhaeoNet. **(B)** A Multi-Dimensional Scaling (MDS) plot of PhaeoNet. The dots correspond to genes and the colors correspond to the WGCNA modules. The tips of the plot correspond to hub genes of PhaeoNet. **(C)** Heatmap plot showing the TOM supplemented by the WGCNA module colors prior to merging. **(D)** Gene dendrogram of all incorporated PhaeoNet genes, obtained by average linkage hierarchical clustering. The first color row underneath the dendrogram shows the WGCNA module assignment obtained by the Dynamic Tree Cut method. The bottom color row shows the merged modules based on a correlation threshold of 0.75.

sheet 2; and **Supplementary Figure 3**), in accordance with other WGCNA studies (Langfelder and Horvath, 2008; Zhao et al., 2010) and following validation by cross-referencing to independently derived gene co-regulation datasets for *P. tricornutum* (described below). The final version of PhaeoNet showed good overall cohesion within the merged modules, as inferrable by multi-dimensional-scaling projection (**Figure 1C**) and correlation heatmaps of gene co-expression interconnectedness (**Figure 1B**).

We present an exemplar merged module output (paleturquoise) in **Figure 2**. A density heatmap, divided vertically by condition and horizontally by gene expression

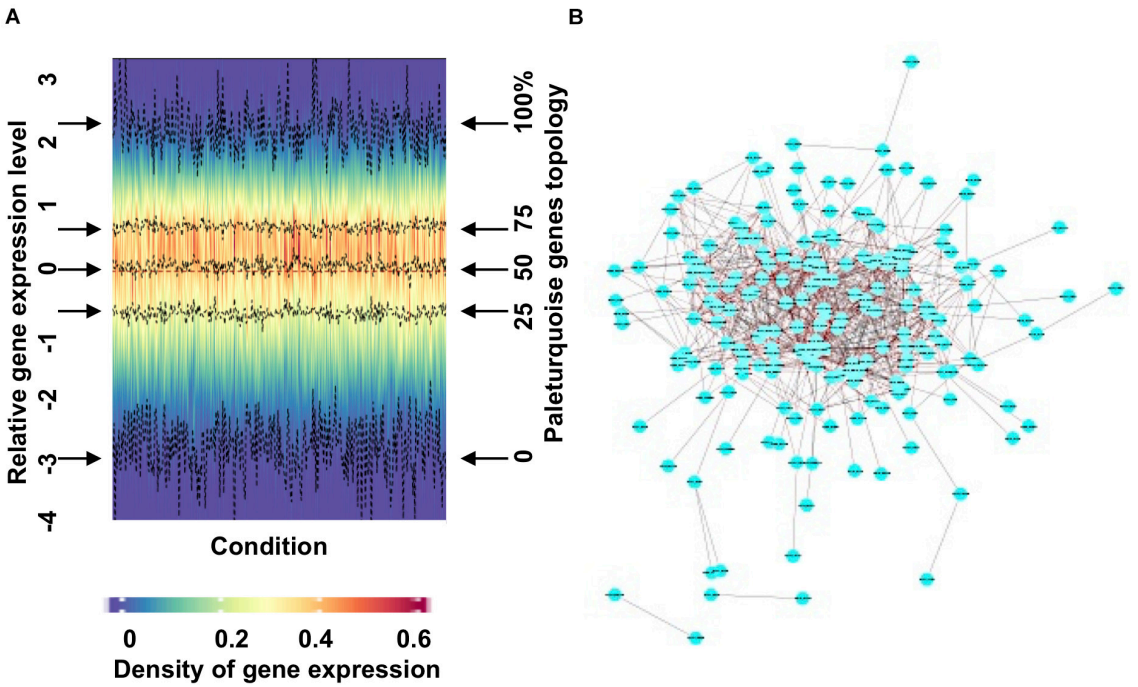


FIGURE 2 | Visualization and analysis of an exemplar PhaeoNet merged module (paleturquoise). **(A)** Density heatmap of all genes assigned to the paleturquoise merged module. The y-coordinate positions in the graph relate to density distribution of gene expression in each sample (shown on the left *y*-axis); the four middle dashed lines (indicated by horizontal arrows on either side of the graph) correspond to the median, first and third quantiles (shown on the right *y*-axis). The majority of the genes in this specific module show limited variation in expression profiles over different conditions samples. **(B)** A topological representation of connectedness within the paleturquoise merged module, visualized with Cytoscape (version 3.6.1). Each node represents a gene. Edges represent pairwise correlations between genes. The network shows all the paleturquoise module genes with a correlation value over a threshold of 0.20.

profile, shows a cohesive module as illustrated by stable values of first quantile, median, and third quantile values (**Figure 2A**) and is defined by high levels of expression across the majority of the conditions explored (**Supplementary Figure 3B**). Cytoscape (Shannon et al., 2003) visualization of the network with a correlation threshold of 0.2 (**Figure 2B**) demonstrates that the paleturquoise merged module is highly connected, showing a cluster of hub genes with high connectivity located in the central part of the network and only a small number of genes with limited connectivity. We provide detailed expression and Cytoscape data for each PhaeoNet merged module via https://osf.io/42xmp.

PhaeoNet Merged Modules Show Concordance With Microarray Co-regulation Data

We tested the reproducibility of our assignations, which may be considered as an independent measure of their robustness, by comparing the repartition of all *P. tricornutum* genes assigned to a PhaeoNet merged module with their corresponding distributions in 500 co-regulated clusters previously assembled from microarray data within the DiatomPortal server (Ashworth et al., 2016; **Supplementary Figure 4A** and **Supplementary Table 2**). Across 7,751 assessable genes with both PhaeoNet and DiatomPortal assignations, we identified 4,127 (53%) that occurred in the same PhaeoNet merged module as another gene with the same DiatomPortal cluster assignation; and 2,751 genes (35%) that occurred within the single PhaeoNet merged module incorporating the greatest number of genes from the same DiatomPortal cluster. Both of these frequencies were judged to be significantly greater than expected through a random distribution ($P = 0$, one-tailed chi-squared test), suggesting strong concordance between both datasets.

From the 461 (83%) DiatomPortal clusters for which we could identify corresponding PhaeoNet merged modules, 369 (80%) were preferentially distributed in one PhaeoNet merged module only, with the greatest number of clusters associated with the darkgray merged module (79 clusters), blue (46 clusters) and cyan (44 clusters) merged modules, reflecting the greater size of each merged module (**Supplementary Figure 4A** and **Supplementary Table 2**). No DiatomPortal clusters were found to be incorporated preferentially into the bisque4, darkmagenta, greenyellow, gray, lightsteelblue1 and mediumpurple3 PhaeoNet merged modules. It is possible that these merged modules represent transcriptional networks not visualized within DiatomPortal due to the different source datasets, generated using different techniques (e.g., microarray versus RNAseq data, assembled with hierarchical clustering versus WGCNA; Ashworth et al., 2016), which may influence what genes are inferred to be coexpressed using each analysis.

We also verified the number of associations independently found between pairs of genes in DiatomPortal clusters and PhaeoNet modules generated with independent merging thresholds, as an independent test of the appropriateness of our selected 0.75 merging threshold (**Supplementary Figure 4B**). We found greater concordance between DiatomPortal and PhaeoNet modules generated with a 0.75 merging threshold, as in our methodology, than in unmerged WGCNA modules, or modules merged with higher (0.8) or lower (0.7) threshold values (**Supplementary Figure 4B**).

Different PhaeoNet Merged Modules Perform Different Biological Activities in the *P. tricornutum* Cell

Next, we profiled the predominant biological activities associated with each merged module by calculating enrichment scores for different functional, subcellular targeting and evolutionary annotations across the *P. tricornutum* genome (Rastogi et al., 2018; **Figure 3** and **Supplementary Figure 5**). A full set of protein annotations *P. tricornutum*, including PhaeoNet module assignations, inferred functions, predicted localization and inferred evolutionary origin, is provided for user exploration in **Supplementary Table 2**.

We identified seven major subsets of merged modules with different biological properties. The first subset consists of merged modules (blue, lightcyan1, lightsteelblue1 and salmon) associated with the chloroplast [either genes encoding chloroplast-targeted proteins, inferred with ASAFind (Gruber et al., 2015) or HECTAR (Gschloessl et al., 2008), or of inferred red algal origin in a previous BLAST top hit analysis of the *P. tricornutum* genome (Rastogi et al., 2018)]. We included proteins of red algal origin as an independent estimator of chloroplastic origin, as the vast majority of red algal protein in *P. tricornutum* likely derive from the diatom chloroplast endosymbiont (Dorrell et al., 2017) and to allow us to detect chloroplast-associated proteins that elude *in silico* targeting prediction (Nonoyama et al., 2019; Schober et al., 2019). These merged modules were also enriched (as inferred with KEGG analysis (Kanehisa, 2017) in genes related to photosynthesis, carbon-fixation and core biosynthetic pathways (e.g., amino acid and pigment biosynthesis) associated with diatom chloroplasts (**Figure 3** and **Supplementary Figure 5**; Nonoyama et al., 2019). Nearly all of the merged modules within this subset were enriched in activating histone marks (e.g., H3K9Ac and H3K14Ac) and depleted in repressive marks (e.g., H3K9me2 and H3K27me3) in cultures grown under replete media conditions (**Figure 3** and **Supplementary Figure 5**; Veluchamy et al., 2015; Zhao et al., 2020), consistent with high levels of expression. Each of the chloroplast-enriched modules contained enrichments in different KEGG functions (discussed below), although only one of these modules (blue) was enriched in proteins containing at least one KEGG annotation (**Supplementary Figure 5**); and, in any case, all merged modules contain substantial numbers (between 18%, bisque4; and 54%, violet).

A second parallel set of merged modules (floralwhite, magenta, mediumpurple3, and orangered4), which was also found to be enriched in activating histone marks, was enriched in genes encoding mitochondria-targeted proteins (inferred with MitoFates, HECTAR and WolfPSort (Horton et al., 2007; Gschloessl et al., 2008; Fukasawa et al., 2015) and mitochondria-associated functions (e.g., oxidative phosphorylation and pyruvate metabolism; **Figure 3** and **Supplementary Figure 5**). Of note, the paleturquoise merged module was uniquely enriched in genes encoding both chloroplast and mitochondria-targeted proteins, suggesting a probable hub between both organelle functions (**Figure 3**).

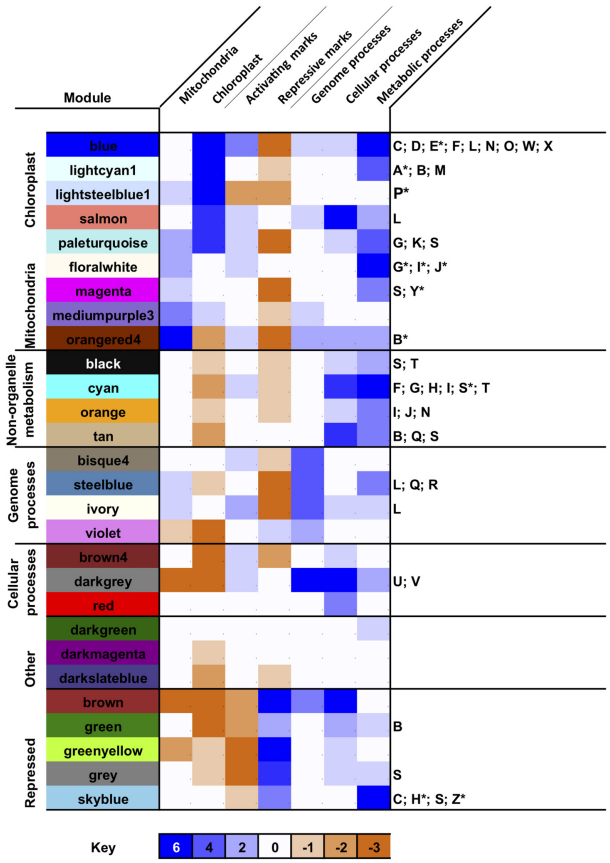


FIGURE 3 | Biological properties associated with PhaeoNet merged modules. This Figure provides an overview of enrichments of different organelle targeting (Horton et al., 2007; Gschloessl et al., 2008; Fukasawa et al., 2015; Gruber et al., 2015), epigenetic (Veluchamy et al., 2013, 2015; Zhao et al., 2020), evolutionary (Rastogi et al., 2018) and KEGG pathway annotations (Kanehisa, 2017) enriched in merged modules. The first seven (shaded) columns provide a score for different conditions, aggregated from chi-squared P-values of multiple enrichment predictors (defined beneath): enrichments in each condition carry a score of +1 if significant to $P < 0.05$ and +2 if significant to $P < 10^{-05}$; and depletions in each condition carry a score of –1 if significant to $P < 0.05$ and –2 if significant to $P < 10^{-05}$, assessed by chi-squared test against a null hypothesis of a random distribution of these features across all genes assigned to a PhaeoNet merged module. The final column lists all metabolic pathways enriched to $P < 0.05$, or $P < 10^{-05}$ (asterisked) for each merged module, assessed by chi-squared test as above. Verbose outputs for each set of conditions are provided in **Supplementary Figure 5**. Additional annotations, e.g., enrichments in inferred evolutionary origins of each merged module, are provided for user exploration in **Supplementary Table S2**.

We identified three further subsets of merged modules that were enriched in cytoplasmic or nuclear processes involved in metabolism (black, cyan, orange, and tan); genome-associated processes including transcription, translation and genome repair (bisque4, steelblue, ivory and violet); or cellular processes including protein modification, protein trafficking and the cell cycle (brown4, darkgray, and red; **Figure 3** and **Supplementary Figure 5**). Certain merged modules contained a mixture of genes encoding both metabolic and non-metabolic proteins: amongst other examples, the steelblue merged module was found to be enriched both in genes encoding proteins associated with ribosome and tRNA biogenesis and also in genes encoding enzymes involved in purine and pyrimidine metabolism, suggesting a probable transcriptional coordination of nucleotide biosynthesis to translational activity in *P. tricornutum* cells (**Figure 3** and **Supplementary Figure 5**). A sixth subset of merged modules (darkgreen, darkmagenta and darkslateblue) showed no obvious enrichment in any KEGG function or organelle localization, except for a possible enrichment in peroxisomal functions in the darkgreen merged module (Davis et al., 2017).

The final merged module subset (brown, green, greenyellow, gray, and skyblue) was uniquely enriched in repressive histone marks and depleted in activating histone marks, in cultures grown on replete media (**Figure 3**; Veluchamy et al., 2015). These merged modules may either be constitutively repressed in *P. tricornutum* cells, or might lose their repressive histone marks and be expressed in alternative conditions to the replete culture conditions in which the epigenetic datasets were collected (Zhao et al., 2020). We noted that the greenyellow merged module, for example, was enriched in proteins with at least one KEGG annotation; and the skyblue merged module was found to be significantly enriched in genes encoding proteins involved in carbon fixation, the TCA cycle and propionate metabolism (**Figure 3** and **Supplementary Figure 5**). Further studies of the epigenetic marks associated with these modules, including under physiological conditions in which they are most highly expressed (**Supplementary Figure 3**) will be necessary to determine under what circumstances the genes they contain make significant contributions to *P. tricornutum* biology.

PhaeoNet Merged Modules Reveal Transcriptional Co-regulation in *P. tricornutum* Chloroplast and Mitochondrial Metabolism

Having noticed specific biases in the distribution of mitochondria- and chloroplast-targeted proteins within our dataset and given the distinctive organelle metabolism noted in diatoms compared to plants (Kroth et al., 2008; Nonoyama et al., 2019; Smith et al., 2019), we wished to identify which key chloroplast and mitochondrial functions are revealed by PhaeoNet to be transcriptionally coordinated with one another. We searched the distribution of 372 manually curated nuclear-encoded proteins with known chloroplast- and mitochondria-associated functions and localizations (**Figure 4**, **Supplementary Figure 6**, and **Supplementary Table 3**). At least one gene encoding one such protein of each merged module was present in this set, however, only 12 merged modules contained more than 10 genes and amounted to 83% of the set.

The most abundantly represented merged module (blue, 69 chloroplast or mitochondrial occurrences) was clearly associated with genes encoding chloroplast anabolic reactions, containing enzymes associated with the Calvin–Benson–Bassham (CBB) cycle, chloroplast-targeted glycolysis/gluconeogenesis (Kroth et al., 2008) and fatty acid synthesis (Maréchal and Lupette, 2020), along with theta-class carbonic anhydrases that mediate biophysical carbon concentrating mechanisms in diatom chloroplasts (**Figure 4** and **Supplementary Figure 6**; Kikutani et al., 2016; Nonoyama et al., 2019). The blue merged module additionally contained genes encoding chloroplast-targeted proteins implicated in photoprotection, including the diatom xanthophyll cycle (e.g., Phatr3_J51703 encoding violaxanthin de-epoxidase; Frommolt et al., 2008; Dautermann and Lohr, 2017), tocopherol synthesis (e.g., Phatr3_J20470, encoding tocopherol cyclase; Dłużewska et al., 2016; Nonoyama et al., 2019) and two genes (Phatr3_J27278 and Phatr3_J44733) encoding LhcX-class chlorophyll-binding proteins, associated with high- and low-light adaptation responses in diatoms (**Supplementary Tables 2**, **3**; Taddei et al., 2016; Buck et al., 2019).

Genes encoding photosynthetic metabolism enzymes were concentrated in the lightcyan1 (41 occurrences) and lightsteelblue1 merged modules (19 occurrences). The lightcyan1 merged module included genes encoding LhcF-, LhcR-, and chlorophyll a/b-binding proteins, which are typically considered not to be involved in light stress responses (Gundermann et al., 2013; Büchel, 2015) and nucleus-encoded subunits of photosystems I, II and cytochrome c_6 (Grouneva et al., 2011; Roncel et al., 2016); whereas the lightsteeblue1 merged module contained the majority of genes involved in diatom chlorophyll and isoprenoid synthesis (Bertrand, 2010; Cihlar et al., 2016). We noted the presence of two genes encoding enzymes involved in pigment biosynthesis, respectively carotenoids (Phatr3_J21829, encoding 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase) and chlorophyll (Phatr3_J30690, encoding 3,8-divinyl protochlorophyllide-a 8-vinyl reductase (Wang et al., 2010) in the lightcyan1 merged module (**Supplementary Figure 6**). We also noted the presence of the gene Phatr3_J47674 encoding the iron stress-induced protein ISIP3 within the lightcyan1 merged module, which may point to a functional role for this protein in chloroplast photosystem assembly (**Supplementary Figure 6**; Allen et al., 2008; Chappell et al., 2015).

Genes encoding mitochondrial respiratory chain proteins were concentrated toward the orangered4 merged module (27 occurrences), whereas, genes encoding TCA cycle enzymes were concentrated toward the cyan merged module (34 occurrences). The orangered4 merged module also contained large numbers of genes encoding mitochondrial ribosomal proteins, which may relate to redox-state dependent regulation of mitochondrial biogenesis pathways (Allen, 2003). In contrast, most genes encoding chloroplast biogenesis-related proteins were identified in separate PhaeoNet merged modules to genes encoding proteins of the photosystem core, with significant enrichments

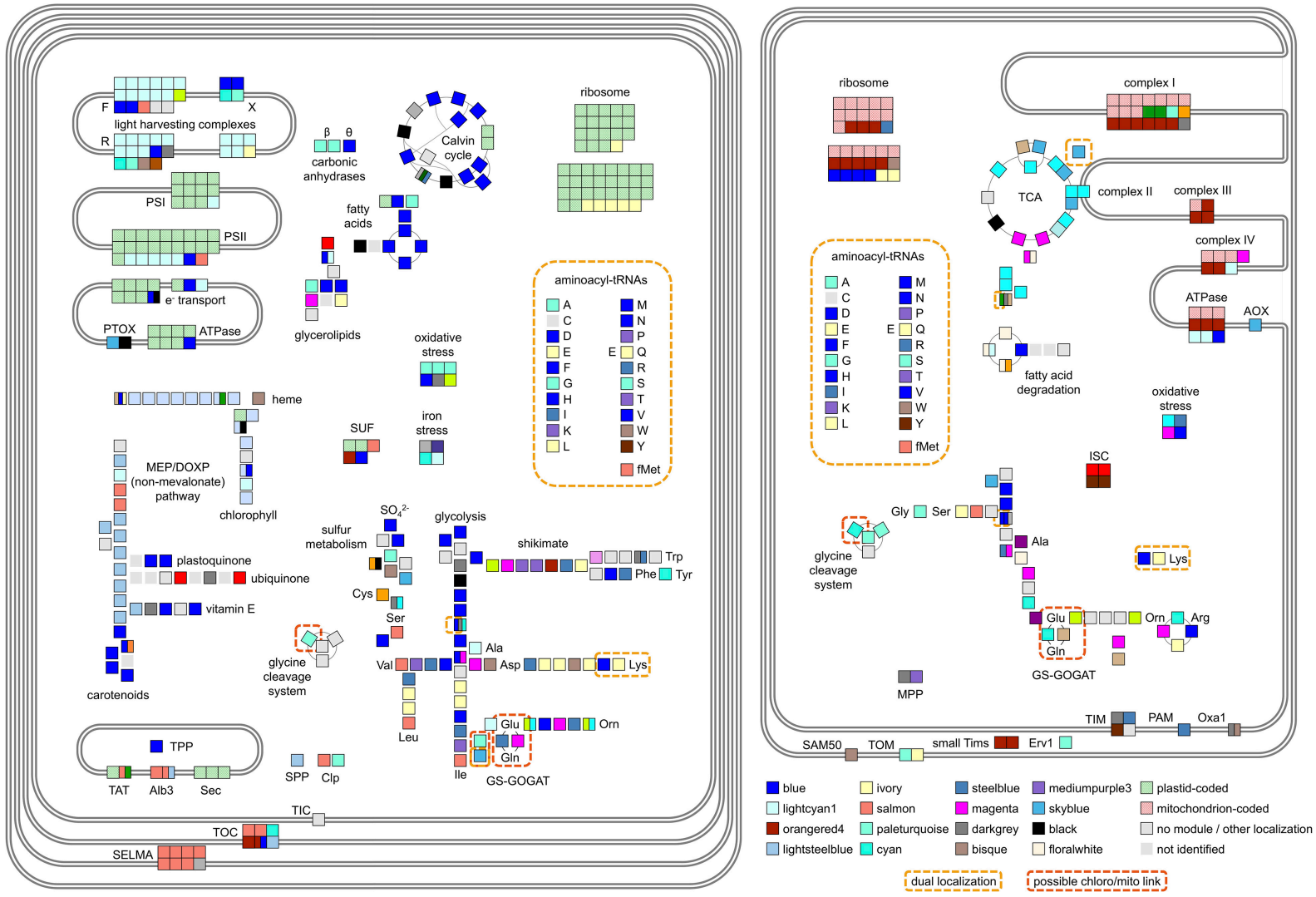


FIGURE 4 | Main metabolic pathways and functional complexes of *P. tricornutum* plastid **(left)** and mitochondrion **(right)** and their composition in regard to PhaeoNet merged modules. Each square represents a gene encoding a protein identified either from N-terminal targeting predictions to function in the chloroplast or mitochondrion. Clusters of adjacent squares pertain to genes encoding different components of a specific multi-unit enzyme or complex; and split squares pertain to genes encoding functional homologs of one specific protein. The assigned merged modules are indicated as their respective colors, with the 16 most abundant merged modules shown in the legend. Additionally, proteins coded in organellar genomes (Oudot-Le Secq et al., 2007; Oudot-Le Secq and Green, 2011; Yu et al., 2018) are shown as dotted green or red; proteins for which chloroplast- or mitochondria-targeted isoforms or merged modules could not be assigned are shown as light gray; and enzymatic steps not identified in the genome are shown as light gray squares without borders. Dual-localized proteins (Gile et al., 2015; Dorrell et al., 2017) are marked by checkered yellow boxes; while orange boxes highlight potential connection points between the two organelles. Abbreviations are as follows: CAs, carbonic anhydrases; MEP/DOXP, mevalonate and non-mevalonate pathways for isoprenoid biosynthesis; SUF, iron-sulfur complex assembly; MPP,/TPP/SPP, mitochondrial, thylakoid and stromal processing peptidases; TAT, twin-arginine-dependent thylakoid protein import pathways; AOX/PTOX, mitochondrial and chloroplast alternative oxidases; TCA, Citric Acid cycle; Orn, ornithine; GCS, glycine shuttle; GS-GOGAT, glutamine synthetase/glutamate synthase shuttle. Detailed enzyme distributions for each pathway are shown in **Supplementary Figure 6**.

of genes encoding chloroplast ribosomal proteins in the ivory merged module (otherwise enriched in chloroplast branched-chain amino acid and lysine biosynthesis (Bromke, 2013). It remains to be determined to what extent the expression of the chloroplast- and mitochondrial-genomes of *P. tricornutum* are regulated in response to the redox state, versus metabolic fluxes experienced in both organelles.

Finally, we considered the repartition of functionally uncharacterized, but conserved domains across chloroplast-targeted proteins in our dataset, focusing on DUFs (Domains of Unknown Function). We found 30 chloroplast-targeted proteins containing at least one DUF and 8 DUFs assigned to at least two chloroplast-targeted proteins (**Supplementary Table 3**, Sheet 2). Amongst these recurrent chloroplast-associated DUFs were two examples (DUF1995 and DUF3493), which have previously been implicated to function in photosystem assembly within thylakoid membranes (Chi et al., 2012; Bohne et al., 2016; Li et al., 2019). Both of these DUFs were found amongst chloroplast-targeted proteins in the paleturquoise merged module (Phatr3_J38149, Phatr3_J40136, and Phatr3_J46926, containing DUF1995; and Phatr3_EG02444, containing DUF3493); in the blue module (Phatr3_J44212, containing DUF1995; and Phatr3_J45569, containing DUF3494); and DUF1995 furthermore occurred in chloroplast-targeted proteins in the lightcyan1 (Phatr3_J44529) and lightsteelblue (Phatr3_J40199) modules (**Supplementary Table 3**, Sheet 2). Each of these modules are enriched in different chloroplast-targeted metabolism pathways (**Figures 3**, **4**), suggesting complex connections between the regulation of chloroplast anabolism and photosystem assembly.

Amongst the other DUFs associated with more than one chloroplast-targeted protein were DUF814, which is implicated in RNA quality control and amongst *P. tricornutum* chloroplast-targeted protein includes one (Phatr3_J45207, within the paleturquoise module) with some structural homology to a ferrous iron transporter (Maxwell Burroughs and Aravind, 2014); and DUF563, which contains a carbohydrate-active domain (Park et al., 2010) and includes at least one chloroplast-targeted protein (Phatr3_EG00581) within the blue module, otherwise implicated in chloroplast carbon metabolism (**Figures 3**, **4**). It remains to be determined if either of these proteins has novel functions, e.g., respectively in iron status sensing or in the diversification of carbohydrate metabolism in the *P. tricornutum* chloroplast, via the generating and phenotyping of mutant lines.

PhaeoNet Merged Modules Identify Complex Crosstalk between the Chloroplast and Mitochondrion in *P. tricornutum*

Previously, intricate metabolic connections have been observed between *P. tricornutum* chloroplasts and mitochondria, which are distinctive to those found in plants (Prihoda et al., 2012; Bailleul et al., 2015; Broddrick et al., 2019; Murik et al., 2019). We wished to determine which of these connections were visible within our data, noting multiple, transcriptionally independent connections between the predicted proteomes of chloroplasts and mitochondria in PhaeoNet data (highlighted in **Supplementary Figure 6**). These included the presence of genes encoding chloroplast-targeted protein import subunits (e.g., Phatr3_J32195 encoding Tic20, Phatr3_EG02421 encoding Tic21) within the otherwise predominantly mitochondrial orangered4 merged module and the presence of large numbers of amino-acyl tRNA synthetase genes (which are typically dual-targeted to the chloroplasts and mitochondria in diatoms (Gile et al., 2015; Dorrell et al., 2017, 2019) in the otherwise chloroplast-associated blue merged module.

We furthermore noted the presence of multiple chloroplast-targeted proteins associated with chloroplast division (e.g., Phatr3_J34093, Phatr3_J42361, and Phatr3_J14995, encoding FtsZ-type division proteins) in the blue module, potentially linking the synthesis of chloroplast and mitochondrial tRNAs to chloroplast replication. A further two proteins implicated in chloroplast replication (e.g., Phatr3_J21455, encoding a dynamin-related DRPB85-class protein and Phatr3_J14426, encoding a further FtsZ protein)—were found in the darkgray module, which was also populated by proteins involved in mitochondrial protein import (MPP, TIM, OXA1; **Supplementary Table 3**), suggesting probable links between chloroplast and mitochondrial biogenesis. Of note, at least two of the FtsZ proteins (Phatr3_J34093, within the blue module and Phatr3_J14426, within the darkgray module) were inferred to possess both chloroplast and mitochondrial-targeting sequences, underpinning the likely coordination of biogenesis of both organelles (**Supplementary Table 3**). This coordination may underpin the close topological associations and synchronized division cycles observed between the *P. tricornutum* mitochondrion and chloroplast observed *in vivo* (Tanaka et al., 2015; Dorrell and Bowler, 2017).

Alongside these more general links, we identified specific points of co-regulation between each organelle. The paleturquoise merged module, as the only merged module found to be enriched in both chloroplast and mitochondria functions (**Figure 3**) was of particular interest and contained genes encoding enzymes participating in several different chloroplast and mitochondria metabolic pathways. These included genes for mitochondria-targeted glycine dehydrogenase (Phatr3_J22187) and serine hydroxymethyltransferase (Phatr3_J32847) and a gene for a chloroplast-targeted dihydrolipoamide dehydrogenase (Phatr3_J30113), which participate (as part of the glycine shuttle) in metabolic recycling of 2-P-glycolate produced through photosynthesis (**Supplementary Figure 6**; Zheng et al., 2013; Davis et al., 2017). The paleturquoise merged module additionally contains a gene encoding mitochondria-targeted malate dehydrogenase (Phatr3_J54082), which may additionally participate in the photorespiratory metabolism of glycolate by allowing the recycling of mitochondrial serine (via pyruvate) in the TCA cycle (Davis et al., 2017; Broddrick et al., 2019). Genes encoding at least three further plastidial oxidative stress-related proteins (Phatr3_J12583, encoding Fe-Mn family superoxide dismutase; Phatr3_J45252, encoding a plastidial thioredoxin; and Phatr3_J31436, encoding a plastidial ortholog of peroxisomal membrane protein 2, Davis et al., 2017; Dorrell et al., 2017) belong to the paleturquoise merged module, underlining its importance in oxidative stress responses.

Genes encoding both glutamine synthase (GS) and glutamate synthase/glutamine oxoglutarate aminotransferase (GOGAT), which have distinct plastidial and mitochondrial homologs in *P. tricornutum* (Broddrick et al., 2019; Smith et al., 2019), belong to different PhaeoNet merged modules (cyan, tan, steelblue and magenta), suggesting a relatively complex regulation of this hub. The plastid-localized GS (encoded by Phatr3_J51092) belongs to the magenta merged module, which also contains the subsequent genes encoding enzymes mediating the entry of GS-produced NH_3 into the mitochondrial ornithine-urea cycle (Phatr3_J42398 encoding malate dehydrogenase; Phatr3_J30145 encoding citrate synthase; Phatr3_J22913 encoding pyruvate kinase), suggesting this co-regulated pathway may have roles in recycling excess NH_3 produced in the chloroplast, in accordance with previous studies (Levering et al., 2016; Broddrick et al., 2019; Smith et al., 2019).

Finally, we noted the presence of genes encoding chloroplast-targeted plastoquinol terminal oxidase (Phatr3_J4283) and mitochondria-targeted alternative oxidase (Phatr3_EG02359), which are both associated with the photoprotective removal of excess metabolic reducing potential in the skyblue merged module (Bailleul et al., 2015; Murik et al., 2019). This merged module, as discussed above, contains genes encoding three successive enzymes associated with the TCA cycle (Phatr3_J40430 encoding α-ketoglutaryl dehydrogenase; Phatr3_J42015 encoding succinyl-CoA synthetase and Phatr3_J41812 encoding succinate dehydrogenase; Kroth et al., 2008), along with methylmalonyl-CoA mutase (Phatr3_J51830), which may allow excess succinyl-CoA to be diverted into lipid synthesis via propionyl-CoA (Helliwell et al., 2011; Valenzuela et al., 2012). This co-regulation underlines the importance of the succinate hub, and presumably both the glyoxylate cycle and ornithine shunt (as sources of mitochondrial α-ketoglutarate), as routes for the mitochondrial dissipation of excess chloroplast reducing potential (Bailleul et al., 2015; Broddrick et al., 2019).

Transcriptional Regulators of Chloroplast-Targeted Proteins Show Separate Expression Dynamics, Informed by Evolutionary History

Finally, given the complex transcriptional partitioning of genes encoding components of chloroplast and mitochondrial metabolism pathways across PhaeoNet data, we investigated what transcriptional drivers might be implicated in the co-regulation of different metabolism-enriched pathway clusters. First, we considered the repartition of a manually curated list of genes encoding proteins implied in histone and transcription-related processes (including transcription factors, TFs; Rayko et al., 2010; Banerjee et al., 2016) across all merged modules (**Supplementary Figure 7** and **Supplementary Table 4**). These genes were most frequently observed (>5% of total merged module genes) in the darkgray, brown and steelblue merged modules (**Supplementary Figure 7** and **Figure 3**). The brown and darkgray merged modules were additionally enriched in KEGG merged modules related to cytoskeleton proteins (**Supplementary Figure 4**), pointing to close links between cytoskeletal organization and transcriptional regulation in diatoms (for example, within organization of the cell cycle (Huysman et al., 2013; Tanaka et al., 2015). The single most abundant TF family, heat shock factor family (HSF) proteins (Rayko et al., 2010), were most frequently detected in the brown, brown4, cyan and skyblue merged modules (> 5 HSFs each, **Supplementary Figure 7**). Notably, both the brown and brown4 merged modules are also enriched in KEGG functions associated with stress responses (protein ubiquitinylation, autophagy and membrane trafficking) (**Supplementary Figure 5**), consistent with previously inferred functions of specific *P. tricornutum* HSFs in the maintenance of cellular fitness (Chen et al., 2014; Egue et al., 2015).

We also found specific repartitions of genes encoding proteins implicated in light- and circadian-dependent transcriptional responses in *P. tricornutum*, e.g., aureochromes and cryptochromes (Takahashi et al., 2007; Banerjee et al., 2016). These proteins typically have cytoplasmic localizations, but through the perception of light and translocation to the nucleus can regulate the expression of core chloroplast metabolic pathways (Kroth et al., 2017). The circadian-regulated Aureochrome 1c (Phatr3_J12346; Banerjee et al., 2016; Kroth et al., 2017) and a cryptochrome-like blue light receptor (Phatr3_J34592) were both found in the blue merged module, implicated in anabolic metabolism; and the light-regulated Aureochrome 1b (Phatr3_J15977) and the blue-light-dependent protochlorophyllide reductase 1 (Phatr3_J12155; Hunsperger et al., 2016; Mann et al., 2017) were both found in the lightsteelblue1 merged module, alongside the majority of genes encoding other pigment biosynthesis enzymes. In contrast, the gene encoding Aureochrome 1a (Phatr3_J49116), which is essential for high light acclimation but appears to be under exclusively circadian (light-independent) regulation, falls within the lightcyan1 merged module of core photosystem-associated genes (**Supplementary Table 4** and **Supplementary Figure 7**; Banerjee et al., 2016; Mann et al., 2017); while RITMO1 (Phatr3_J44962), associated with the *P. tricornutum* circadian clock, falls within the skyblue merged module, which contains limited chloroplast-related functions except for alternative electron flow pathways (**Supplementary Table 4** and **Supplementary Figure 7**; Annunziata et al., 2019). The separate distributions of light- and circadian-regulated chloroplast regulators might reflect a circadian-entrained synthesis of the core photosynthetic machinery (via Aureochrome 1a), independent of light status, with chloroplast biosynthesis pathways upregulated both by circadian signaling (via Aureochrome 1c) and as a function of light availability (via Aureochrome 1b). This is reminiscent of circadian gene expression patterns visualized in plant and other algal lineages (e.g., the green alga *Ostreococcus* and the dinoflagellate *Lingulodinium*), in which photosynthesis and plastid biogenesis proteins are either expressed at separate times of the day, or show different regulatory responses to circadian and light signals (Wang et al., 2005; Monnier et al., 2010; Noordally et al., 2013). Finally, the gene encoding the Aureochrome 2 protein (Phatr3_J8113), which lacks the conserved flavin-binding domain required for light perception (Takahashi et al., 2007; Kroth et al., 2017), falls within the greenyellow merged module of generally transcriptionally

repressed proteins (**Figure 3**), underlining its independence of chloroplast functions.

Finally, we wished to consider within our dataset what transcriptional dynamics within the nuclear genome may underpin chloroplast gene expression in *P. tricornutum*. Chloroplast transcription in *P. tricornutum*, as in other diatoms, is performed by a plastid-encoded RNA polymerase, unlike the situation in plants in which both plastid- and nuclear-encoded and plastid-targeted polymerases participate (Oudot-Le Secq et al., 2007; Yu et al., 2018). Plastid-encoded RNA polymerases in plants typically interact with nucleus-encoded sigma factors, which may direct them to specific target genes, in response to different regulatory and physiological signals (Shimizu et al., 2010; Noordally et al., 2013). Eight genes are annotated in the *P. tricornutum* nuclear genome to encode sigma factor related proteins (Rayko et al., 2010; **Supplementary Table 4**), but the functions of each protein with regard to the expression of the chloroplast genome remain unclear.

We investigated the functions of *P. tricornutum* sigma factors by combining the repartition of each sigma factor in PhaeoNet with predicted *in silico* localizations of *P. tricornutum* proteins and their closest homologs from other diatom species, as resolved with a single-gene (RAxML) tree (**Figure 5**). Three of the sigma factor genes in *P. tricornutum* possess chloroplast-targeting sequences, as inferred by *in silico* prediction with HECTAR and ASAFind (Gschloessl et al., 2008; Gruber et al., 2015). One of these proteins (Phatr3_J14599, SIGMA1a) falls within the paleturquoise merged module, which is otherwise enriched in chloroplast-related functions pertaining to carbon concentration and the glycine shunt (**Figures 3–5**); while the two remaining chloroplast-targeted proteins (Phatr3_J3388, SIGMA1b; Phatr3_J17029, SIGMA3) fall within the steelblue module, which otherwise lacks obvious enrichments in chloroplast-targeted functions and instead seems to be most closely connected to nucleotide metabolism (**Figures 3**, **5**). Phylogenetic analysis of these three sigma factors indicate that many of their closest

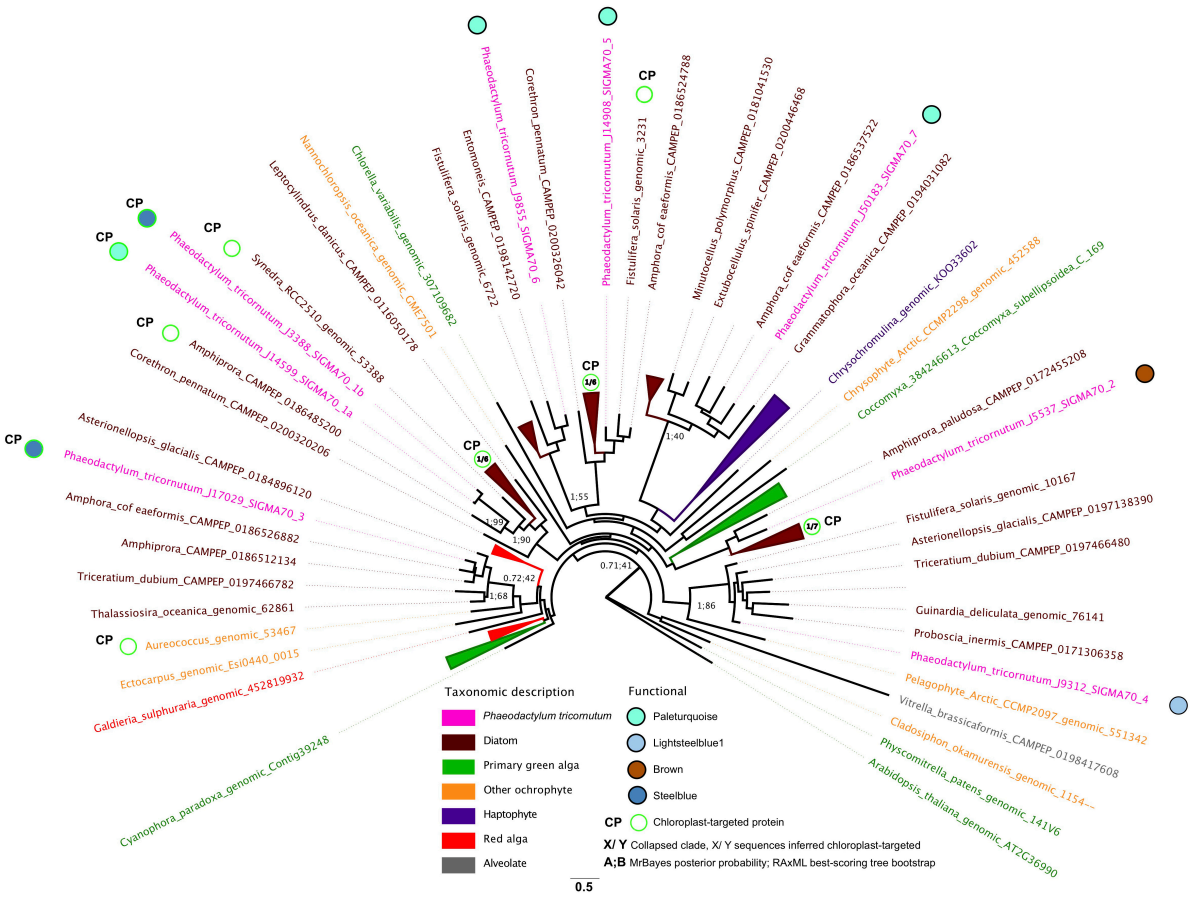


FIGURE 5 | Phylogenetic and transcriptional dynamics of *P. tricornutum* sigma factors. This Figure shows an unrooted best-scoring tree topology for an 86 taxa x 453 aa alignment of subsampled diatom and non-diatom sigma factors and realized using MrBayes v 3.2.7a with the Jones substitution matrix, 600,000 generations, two start chains and 0.5 burnin thresholds (Huelsenbeck and Ronquist, 2001); and RAxML v 8.2 with the PROTGAMMAJTT substitution model with 300 bootstrap replicates (Stamatakis, 2014). Chloroplast-targeting predictions were performed using ASAFind with SignalP v 3.0 (Gruber et al., 2015); and HECTAR (Gschloessl et al., 2008) under default conditions. Branches are colored by phylogenetic affiliation and bootstrap values of nodes recovered with > 40% support are shown. Eight *P. tricornutum* sigma factors are labeled with PhaeoNet merged module repartition and chloroplast targeting sequences were predicted by HECTAR or ASAFind (Gruber et al., 2015).

relatives are sequences with chloroplast-targeting signals from other diatoms, and indeed SIGMA1a and SIGMA1b appear to be recently derived paralogs of one another (**Figure 5**), indicating that they are likely to be conserved parts of the diatom chloroplast transcriptional machinery. The repartition of SIGMA1b and SIGMA3 within a transcriptional module that is largely related to non-chloroplast processes may allow hierarchical control of chloroplast transcription in response to non-chloroplast signals in *P. tricornutum* (e.g., coordination with circadian or cell cycles, Noordally et al., 2013; Tanaka et al., 2015).

The remaining five *P. tricornutum* sigma factors were not predicted to be targeted to the chloroplast and phylogenetic analysis indicated that their closest diatom relatives primarily also lacked chloroplast-targeting signals (**Figure 5**). One of these non-chloroplast-associated sigma factors (Phatr3_J5537, SIGMA2) fell within the largely chloroplast-independent brown module, suggesting that it has non-chloroplastic functions. In contrast, the remaining non-chloroplast targeted sigma factors fell within modules otherwise enriched in chloroplast-associated functions; either lightsteelblue (Phatr3_J9312, SIGMA4), or paleturquoise (Phatr3_J14908, Phatr3_J9855, Phatr3_J50183; SIGMA 5-7; **Figures 3**, **5**). It remains to be determined whether these sigma factors are targeted to the *P. tricornutum* chloroplast, but using alternative methods to those recognized by HECTAR or ASAFind, as per certain other diatom proteins (Kazamia et al., 2018; Schober et al., 2019); function in compartments other than the chloroplast, but participate indirectly in the regulation, e.g., of nucleus-encoded proteins implicated in chloroplast metabolism; or have functions independent of the chloroplast, as has been documented for some other eukaryotic sigma factors (Shadel and Clayton, 1995; Beardslee et al., 2002). These different possibilities may be best discriminated by the experimental characterization, e.g., through mutagenesis and functional phenotyping, of individual *P. tricornutum* sigma factor genes.

CONCLUDING REMARKS

In this project, we have used WGCNA to build an integrated network of *P. tricornutum* gene co-regulation, which we name "PhaeoNet." Our model is able to retrieve well established biological pathways (e.g., chloroplast photosynthetic, anabolic metabolism; and mitochondrial respiration, **Figure 4**) and compares favorably to existing (e.g., microarray-based; Ashworth et al., 2016) studies of gene co-regulation for this species (Ashworth et al., 2016; **Figures 1**, **2** and **Supplementary Figures 1–4**). Moreover, our dataset carries the advantage of decomposing the *P. tricornutum* genome into a smaller number (28) of functionally distinct modules than produced by DiatomPortal. We have integrated these data into previously generated functional, targeting and evolutionary analyses of the *P. tricornutum* genome, allowing us to gain holistic insights into the processes underpinning the gene co-regulation of specific biological processes and organelle metabolic pathways pertinent to diatom biology (**Figure 3** and **Supplementary Figure 5**).

Through a deeper inspection of genes encoding chloroplast and mitochondria-targeted proteins within these data, we identify PhaeoNet merged modules underpinning anabolic (blue), photosynthetic (lightsteelblue 1, lightcyan1) and respiratory (orangered4, cyan) metabolism, and identify multiple metabolic connections between the chloroplast and mitochondria. These include the glycine shunt within the paleturquoise merged module; the ornithine-urea cycle within the magenta merged module; and coordinated chloroplast and mitochondrial alternative oxidase activities in the skyblue merged module; **Figure 4** and **Supplementary Figure 6**. Finally, considering the repartition of transcription-related proteins within our data, we identify probable cognate regulators for different co-ordinated metabolic pathways (**Figure 5** and **Supplementary Figure 7**), demonstrating different associations of aureochrome transcription factors with different chloroplast metabolic pathways. We notably identify hidden diversity in the range of sigma factor genes in the *P. tricornutum* genome, some of which are likely to be involved in the transcriptional regulation of different chloroplast-encoded genes in response to different physiological signals, while others are likely to have different functions to chloroplast gene expression.

The PhaeoNet dataset may be usable as a predictive tool for the characterization of poorly understood proteins, either directly in *P. tricornutum*, as a well-studied model diatom species, or in other diatom or microalgal species for which homologs of *P. tricornutum* proteins are known either from genome or transcriptome datasets (e.g., Keeling et al., 2014; Carradec et al., 2018; Sato et al., 2020). We stress that biological processes elucidated in this species may not necessarily be directly extrapolatable to other algal species; and examples are already known of proteins (e.g., proteins involved in iron-stress tolerance and C4 photosynthesis) that may have different physiological functions even between different diatoms (Kustka et al., 2014; Lampe et al., 2018). Cross-comparisons between PhaeoNet and other data, e.g., gene coregulation datasets erected in other, less well-studied species (Ashworth et al., 2016; Ashworth and Ralph, 2018); environmental expression trends (Carradec et al., 2018); and the phenotypes of a wider range of mutant lines generated in *P. tricornutum* will be essential to understanding the diversity of functions performed by understudied proteins in diatoms and other algae. Nonetheless, insights from our data, delivering actors and signatures of metabolic co-regulation in diatoms, will provide a useful community resource for subsequent directed experimental investigation.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://osf.io/42xmp/.

AUTHOR CONTRIBUTIONS

OA-M was responsible for the design and construction of PhaeoNet. AMGNV and NJ performed the functional analysis of the PhaeoNet modules. YL and XZ participated in the construction of the data used for functional analysis. AG, LT, and CB were responsible for the supervision of the construction of

PhaeoNet. RGD was responsible for the supervision of functional analysis. OA-M and RGD wrote the manuscript, with input from all other co-authors. All authors contributed to the article and approved the submitted version.

FUNDING

RGD and AMGNV acknowledge funding from a CNRS Momentum Fellowship (awarded to RGD, 2019-2021). CB acknowledges funding from the European Research Council for an Advanced Award (grant ERC 835067-DIATOMIC), grants from the French Agence Nationale de la Recherche (MEMOLIFE, ref. ANR10-LABX-54, OCEANOMICS, ref. ANR-11-BTBR-0008 and BrownCut, ref. ANR-19-CE20-0020) and Research Grant "Green Life in the Dark" (RGP0003/2016) from the Human Frontier Science Program. LT acknowledges funding from the CNRS, the region of Pays de la Loire (ConnecTalent EPIALG project) and Nantes métropoles. XZ acknowledges a Ph.D. fellowship from the Chinese Scholarship Council (CSC-201604910722) and funding from Région Pays de la Loire (Awarded to LT).

ACKNOWLEDGMENTS

We would like to thank Dr. Justin Ashworth (University of Technology Sydney) and Dr. Serdar Turkarslan (Institute for Systems Biology, Seattle) for the kind provision of analogous cluster composition data from the DiatomPortal project and Prof. Angela Falciatore (Institut de Biologie Physico-Chimique, Paris) for assistance with the annotation of *P. tricornutum* transcription factors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.590949/full#supplementary-material

REFERENCES

Aibar, S., Fontanillo, C., Droste, C., and De Las Rivas, J. (2015). Functional gene networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. *Bioinformatics* 31, 1686–1688. doi: 10.1093/bioinformatics/btu864

Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10438–10443. doi: 10.1073/pnas.0711370105

Allen, J. F. (2003). The function of genomes in bioenergetic organelles. *Phil. Trans. R. Soc. Biol.* 358, 19–37. doi: 10.1098/rstb.2002.1191

Annunziata, R., Ritter, A., Fortunato, A. E., Manzotti, A., Cheminant-Navarro, S., Agier, N., et al. (2019). bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom P. *tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 13137–13142. doi: 10.1073/pnas.1819660116

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2019). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*. 36, 2251–2252. doi: 10.1093/bioinformatics/btz859

Ashworth, J., and Ralph, P. J. (2018). An explorable public transcriptomics compendium for eukaryotic microalgae. *bioRxiv[Preprint]*. doi: 10.1101/403063

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*. *Mar. Genom.* 26, 21–28. doi: 10.1016/j.margen.2015.10.011

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., and Tanaka, A. (2015). Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. *Nature* 524, 366–371. doi: 10.1038/nature14599

Banerjee, A., Herman, E., Serif, M., Maestre-Reyna, M., Hepp, S., Pokorny, R., et al. (2016). Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. *Nucl. Acids Res.* 44, 5957–5970. doi: 10.1093/nar/gkw420

Barabási, A. L. (2009). Scale-free networks: a decade and beyond. *Science* 325, 412–413. doi: 10.1126/science.1173299

Bártová, E., Krejčí, J., Harničarová, A., Galiová, G., and Kozubek, S. (2008). Histone modifications and nuclear architecture: a review. *J. Histochem. Cytochem* 56, 711–721. doi: 10.1369/jhc.2008.951251

Beardslee, T. A., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S., Stern, D. B., et al. (2002). A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. *Plant. J.* 31, 199–209. doi: 10.1046/j.1365-313x.2002.01344.x

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol* 340, 783–795. doi: 10.1016/j.jmb.2004.05.028

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: 10.1038/nature07517

Bertrand, M. (2010). Carotenoid biosynthesis in diatoms. *Photosynthesis Res.* 106, 89–102. doi: 10.1007/s11120-010-9589-x

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21. doi: 10.1101/gad.947102

Bohne, A.-V., Schwenkert, S., Grimm, B., and Nickelsen, J. (2016). Roles of tetratricopeptide repeat proteins in biogenesis of the photosynthetic apparatus. *Int. Rev. Cell Mol. Biol.* 324, 187–227. doi: 10.1016/bs.ircmb.2016.01.005

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The P. *tricornutum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Bowler, C., Vardi, A., and Allen, A. E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci.* 2, 333–365. doi: 10.1146/annurev-marine-120308-081051

Broddrick, J. T., Du, N., Smith, S. R., Tsuji, Y., Jallet, D., and Ware, M. A. (2019). Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. *New Phytol.* 222, 1364–1379. doi: 10.1111/nph.15685

Bromke, M. A. (2013). Amino Acid biosynthesis pathways in diatoms. *Metabolites* 3, 294–311. doi: 10.3390/metabo3020294

Büchel, C. (2015). Evolution and function of light harvesting proteins. *J. Plant. Physiol.* 172, 62–75. doi: 10.1016/j.jplph.2014.04.018

Buck, J. M., Sherman, J., Río Bártulos, C., Serif, M., Halder, M., and Henkel, J. (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. *Nat. Commun.* 10:4167. doi: 10.1038/s41467-019-12043-6

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9:373. doi: 10.1038/s41467-017-02342-1

Chappell, P. D., Whitney, L. P., Wallace, J. R., Darer, A. I., Jean-Charles, S., and Jenkins, B. D. (2015). Genetic indicators of iron limitation in wild populations

of Thalassiosira oceanica from the northeast Pacific Ocean. *ISME J.* 9, 592–602. doi: 10.1038/ismej.2014.171

Chen, Z., Yang, M. K., Li, C. Y., Wang, Y., Zhang, J., Wang, D. B., et al. (2014). Phosphoproteomic analysis provides novel insights into stress responses in *Phaeodactylum tricornutum*, a model diatom. *J Proteom. Res.* 13, 2511–2523. doi: 10.1021/pr401290u

Chi, W., Ma, J., and Zhang, L. (2012). Regulatory factors for the assembly of thylakoid membrane protein complexes. *Phil. Trans. R. Soc. B* 367, 3420–3429. doi: 10.1098/rstb.2012.0065

Cihlar, J., Fussy, Z., Horak, A., and Obornik, M. (2016). Evolution of the tetrapyrrole biosynthetic pathway in secondary algae: conservation, redundancy and replacement. *PLoS One* 11:0166338. doi: 10.1371/journal.pone.0166338

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 Phytol.* 210, 497–510. doi: 10.1111/nph.13787

Dautermann, O., and Lohr, M. (2017). A functional zeaxanthin epoxidase from red algae shedding light on the evolution of light-harvesting carotenoids and the xanthophyll cycle in photosynthetic eukaryotes. *Plant J.* 92, 879–891. doi: 10.1111/tpj.13725

Davis, A., Abbriano, R., Smith, S. R., and Hildebrand, M. (2017). Clarification of photorespiratory processes and the role of Malic Enzyme in diatoms. *Protist* 168, 134–153. doi: 10.1016/j.protis.2016.10.005

Dłużewska, J., Szymanìska, R., Gabruk, M., Koìs, P. B., Nowicka, B., and Kruk, J. (2016). Tocopherol cyclases- substrate specificity and phylogenetic relations. *PLoS One* 11:0159629. doi: 10.1371/journal.pone.0159629

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformat.* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dorrell, R. G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., et al. (2019). Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6914–6923. doi: 10.1073/pnas.1819976116

Dorrell, R. G., and Bowler, C. (2017). Secondary plastids of stramenopiles. *Adv. Bot. Res.* 84, 59–103.

Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., et al. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6, 23717. doi: 10.7554/eLife.23717

Egue, F., Chenais, B., Tastard, E., Marchand, J., Hiard, S., Gateau, H., et al. (2015). Expression of the retrotransposons Surcouf and Blackbeard in the marine diatom *Phaeodactylum tricornutum* under thermal stress. *Phycologia* 54, 617-627.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Frommolt, R., Werner, S., Paulsen, H., Goss, R., Wilhelm, C., Zauner, S., et al. (2008). Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol. Biol. Evol.* 25, 2653–2667. doi: 10.1093/molbev/msn206

Fukasawa, Y., Tsuji, J., Fu, S. C., Tomii, K., Horton, P., Imai, K., et al. (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteom.* 14, 1113–1126. doi: 10.1074/mcp.M114.043083

Gasch, A. P., and Eisen, M. B. (2002). Exploring the conditional co-regulation of yeast gene expression through fuzzy k-means clustering. *Genom. Biol.* 3:0059. doi: 10.1186/gb-2002-3-11-research0059

Gile, G. H., Moog, D., Slamovits, C. H., Maier, U. G., and Archibald, J. M. (2015). Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genom. Biol. Evol.* 7, 1728–1742. doi: 10.1093/gbe/evv095

Glasser, G. J., and Winter, R. F. (1961). Critical values of the coefficient of rank correlation for testing the hypothesis of independence. *Biometrika* 48, 444–448. doi: 10.2307/2332767

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. Proteom. Res.* 10, 5338–5353. doi: 10.1021/pr200600f

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. doi: 10.1111/tpj.12734

Gschloessl, B., Guermeur, Y., and Cock, J. M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformat.* 9:393. doi: 10.1186/1471-2105-9-393

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532, 465–470. doi: 10.1038/nature16942

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*. *Biochim. Biophys. Acta.* 1827, 303–310. doi: 10.1016/j.bbabio.2012.10.017

Helliwell, K. E., Wheeler, G. L., Leptos, K. C., Goldstein, R. E., and Smith, A. G. (2011). Insights into the evolution of vitamin B12 auxotrophy from sequenced algal genomes. *Mol. Biol. Evol.* 28, 2921–2933. doi: 10.1093/molbev/msr124

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucl. Acids Res.* 35, 585–587. doi: 10.1093/nar/gkm259

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Hunsperger, H. M., Ford, C. J., Miller, J. S., and Cattolico, R. A. (2016). Differential regulation of duplicate light-dependent protochlorophyllide oxidoreductases in the diatom *Phaeodactylum tricornutum*. *PLoS One* 11:0158614. doi: 10.1371/journal.pone.0158614

Huysman, M. J., Fortunato, A. E., Matthijs, M., Costa, B. S., Vanderhaeghen, R., Van den Daele, H., et al. (2013). AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *Plant Cell* 25, 215–228. doi: 10.1105/tpc.112.106377

Kanehisa, M. (2017). Enzyme Annotation and Metabolic Reconstruction Using KEGG. *Methods Mol. Biol.* 1611, 135–145. doi: 10.1007/978-1-4939-7015-5_11

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. doi: 10.1002/pro.3711

Katoh, K., Misawa, K., Kuma, K.-I., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R. G., Vieira, F. R. J., Mach, J., et al. (2018). Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci. Adv.* 16:5. doi: 10.1126/sciadv.aar4536

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., 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 Biol.* 10:1001889. doi: 10.1073/pnas.1603112113

Kikutani, S., Nakajima, K., Nagasato, C., Tsuji, Y., Miyatake, A., and Matsuda, Y. (2016). Thylakoid luminal theta-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9828–9833. doi: 10.1073/pnas.1603112113

Kim, E.-D., and Sung, S. (2012). Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci.* 17, 16–21. doi: 10.1016/j.tplants.2011.10.008

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., and Parker, M. S. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 3:e1426. doi: 10.1371/journal.pone.0001426

Kroth, P. G., Wilhelm, C., and Kottke, T. (2017). An update on aureochromes: phylogeny - mechanism - function. *J. Plant Physiol.* 217, 20–26. doi: 10.1016/j.jplph.2017.06.010

Kustka, A., Milligan, A. J., Zheng, H., New, A. M., Gates, C., Bidle, K. D., et al. (2014). Low CO2 results in a rearrangement of carbon metabolism to support C4 photosynthetic carbon assimilation in Thalassiosira pseudonana. *New Phytol.* 204, 507–520. doi: 10.1111/nph.12926

Lampe, R. H., Mann, E. L., Cohen, N. R., Till, C. P., Thamatrakoln, K., Brzezinski, M. A., et al. (2018). Different iron storage strategies among bloom-forming

diatoms. *Proc. Natl. Acad. Sci. U.S.A* 115, 12275–12284. doi: 10.1073/pnas.1805243115

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformat.* 9:559. doi: 10.1186/1471-2105-9-559

Levering, J., Broddrick, J., Dupont, C. L., Peers, G., Beeri, K., Mayers, J., et al. (2016). Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One* 11:0155038. doi: 10.1371/journal.pone.0155038

Li, X., Patena, W., Fauser, F., Jinkerson, R. E., Saroussi, S., Meyer, M. T., et al. (2019). A genome-wide algal mutant library reveals a global view of genes required for eukaryotic photosynthesis. *Nat. Genet.* 1, 627–635. doi: 10.1038/s41588-019-0370-6

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. doi: 10.1093/bioinformatics/btt656

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genom. Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Maheswari, U., Jabbari, K., Petit, J.-L., Porcel, B. M., Allen, A. E., Cadoret, J.-P., et al. (2010). Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genom. Biol.* 11:85. doi: 10.1186/gb-2010-11-8-r85

Mann, M., Serif, M., Jakob, T., Kroth, P. G., and Wilhelm, C. (2017). PtAUREO1a and PtAUREO1b knockout mutants of the diatom *Phaeodactylum tricornutum* are blocked in photoacclimation to blue light. *J. Plant Physiol.* 217, 44–48. doi: 10.1016/j.jplph.2017.05.020

Maréchal, E., and Lupette, J. (2020). Relationship between acyl-lipid and sterol metabolisms in diatoms. *Biochimie* 169, 3–11. doi: 10.1016/j.biochi.2019.07.005

Maxwell Burroughs, A., and Aravind, L. (2014). A highly conserved family of domains related to the DNA-glycosylase fold helps predict multiple novel pathways for RNA modifications. *RNA Biol.* 11, 360–372. doi: 10.4161/rna.28302

McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., et al. (2017). Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell* 29, 2047–2070. doi: 10.1105/tpc.16.00910

Michalak, P. (2008). Coexpression, co-regulation and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91, 243–248. doi: 10.1016/j.ygeno.2007.11.002

Monnier, A., Liverani, S., Bouvet, R., Jesson, B., Smith, J. Q., Mosser, J., et al. (2010). Orchestrated transcription of biological processes in the marine picoeukaryote Ostreococcus exposed to light/dark cycles. *BMC Genom* 11:192. doi: 10.1186/1471-2164-11-192

Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J., Couce, M. L., and Cocho, J. A. (2013). A glimpse into past, present and future DNA sequencing. *Mol. Genet. Metab.* 110, 3–24. doi: 10.1016/j.ymgme.2013.04.024

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.* 35, 182–185. doi: 10.1093/nar/gkm321

Murik, O., Tirichine, L., Prihoda, J., Thomas, Y., Araújo, W. L., Allen, A. E., et al. (2019). Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. *New Phytol.* 221, 1303–1316. doi: 10.1111/nph.15479

Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., et al. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules* 9:322. doi: 10.3390/biom9080322

Noordally, Z. B., Ishii, K., Atkins, K. A., Wetherill, S. J., Kusakina, J., Walton, E. J., et al. (2013). Circadian control of chloroplast transcription by a nuclear-encoded timing signal. *Science* 339, 1316–1319. doi: 10.1126/science.1230397

Norbury, C. J. (2010). 3' uridylation and the regulation of RNA function in the cytoplasm. *Biochem. Soc. Trans.* 38, 1150–1153. doi: 10.1042/bst0381150

Novák Vanclová, A. M. G., Zoltner, M., Kelly, S., Soukal, P., Záhonová, K., Füssy, Z., et al. (2020). Metabolic quirks and the colourful history of the Euglena gracilis secondary plastid. *New Phytol.* 225, 1578–1592. doi: 10.1111/nph.16237

Osborn, H. L., and Hook, S. E. (2013). Using transcriptomic profiles in the diatom *Phaeodactylum tricornutum* to identify and prioritize stressors. *Aquat. Toxicol.* 13, 12–25. doi: 10.1016/j.aquatox.2013.04.002

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. doi: 10.1016/j.gene.2011.02.001

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. *Mol. Genet. Genom.* 277, 427–439. doi: 10.1007/s00438-006-0199-4

Park, B. H., Karpinets, T. V., Syed, M. H., Leuze, M. R., and Uberbacher, E. C. (2010). CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20, 1574–1584. doi: 10.1093/glycob/cwq106

Parks, M. B., Nakov, T., Ruck, E. C., Wickett, N. J., and Alverson, A. J. (2018). Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *Am. J. Bot.* 105, 330–347. doi: 10.1002/ajb2.1056

Prihoda, J., Tanaka, A., de Paula, W. B. M., Allen, J. F., Tirichine, L., and Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* 63, 1543–1557. doi: 10.1093/jxb/err441

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., 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. doi: 10.1038/s41598-018-23106-x

Rayko, E., Maumus, F., Maheswari, U., Jabbari, K., and Bowler, C. (2010). Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytol.* 188, 52–66. doi: 10.1111/j.1469-8137.2010.03371.x

Reja, R., Vinayachandran, V., Ghosh, S., and Pugh, B. F. (2015). Molecular mechanisms of ribosomal protein gene co-regulation. *Genes Dev.* 29, 1942–1954. doi: 10.1101/gad.268896.115

Río Bártulos, C., Rogers, M. B., Williams, T. A., Gentekaki, E., Brinkmann, H., Cerff, R., et al. (2018). Mitochondrial glycolysis in a major lineage of eukaryotes. *Genom. Biol. Evol.* 10, 2310–2325. doi: 10.1093/gbe/evy164

Roncel, M., Gonzalez-Rodriguez, A. A., Naranjo, B., Bernal-Bayard, P., Lindahl, A. M., Hervas, M., et al. (2016). Iron deficiency induces a partial inhibition of the photosynthetic electron transport and a high sensitivity to light in the diatom *P. tricornutum*. *Front. Plant Sci.* 7:14. doi: 10.3389/fpls.2016.01050

Sato, S., Nanjappa, D., Dorrell, R. G., Vieira, F. R. J., Kazamia, E., Tirichine, L., et al. (2020). Genome-enabled phylogenetic and functional reconstruction of an araphid pennate diatom Plagiostriata sp. *CCMP*470, previously assigned as a radial centric diatom and its bacterial commensal. *Sci. Rep.* 10:9449. doi: 10.1038/s41598-020-65941-x

Schober, A., Río Bártulos, C., Bischoff, A., Lepetit, B., Gruber, A., and Kroth, P. G. (2019). Organelle studies and proteome analyses of mitochondria and plastids fractions from the diatom Thalassiosira pseudonana. *Plant Cell Physiol.* 60, 1811–1828. . https://doi.org/10.1093/pcp/pcz097, doi: 10.1093/pcp/pcz097

Shadel, G. S., and Clayton, D. A. (1995). A Saccharomyces cerevisiae mitochondrial transcription factor, sc-mtTFB, shares features with sigma factors but is functionally distinct. *Mol. Cell Biol.* 15, 2101–2108. doi: 10.1128/mcb.15.4.2101

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genom. Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shimizu, M., Kato, H., Ogawa, T., Kurachi, A., Nakagawa, Y., and Kobayashi, H. (2010). Sigma factor phosphorylation in the photosynthetic control of photosystem stoichiometry. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10760–10764. doi: 10.1073/pnas.0911692107

Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Oborník, M., Horák, A., et al. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* 10:4552. doi: 10.1038/s41467-019-12407-y

Smith, S. R., Gillard, J. T., Kustka, A. B., McCrow, J. B., Badger, J. H., Zheng, H., et al. (2016). Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation.". *PLoS Genet* 12:e1006490. doi: 10.1371/journal.pgen.1006490

Snel, B., van Noort, V., and Huynen, M. A. (2004). Gene co-regulation is highly conserved in the evolution of eukaryotes and prokaryotes. *Nucl. Acids Res.* 32, 4725–4731. doi: 10.1093/nar/gkh815

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Taddei, L., Stella, G. R., Rogato, A., Bailleul, B., Fortunato, A. E., Annunziata, R., et al. (2016). Multisignal control of expression of the LHCX protein family in the marine diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 67, 3939–3951. doi: 10.1093/jxb/erw198

Takahashi, F., Yamagata, D., Ishikawa, M., Fukamatsu, Y., Ogura, Y., Kasahara, M., et al. (2007). AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19625–19630. doi: 10.1073/pnas.0707692104

Tanaka, A., De Martino, A., Amato, A., Montsant, A., Mathieu, B., Rostaing, P., et al. (2015). Ultrastructure and membrane traffic during cell division in the marine pennate diatom *Phaeodactylum tricornutum*. *Protist* 166, 506–521. doi: 10.1016/j.protis.2015.07.005

Teichmann, S. A., and Babu, M. M. (2002). Conservation of gene co-regulation in prokaryotes and eukaryotes. *Trends Biotechnol.* 20, 407–410. doi: 10.1016/s0167-7799(02)02032-2

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., et al. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693. doi: 10.1126/science.1192002

Valenzuela, J., Mazurie, A., Carlson, R. P., Gerlach, R., Cooksey, K. E., Peyton, B. M., et al. (2012). Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 5:40. doi: 10.1186/1754-6834-5-40

Valle, K. C., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., et al. (2014). System responses to equal doses of photosynthetically usable radiation of blue, green and red light in the marine diatom *Phaeodactylum tricornutum*. *PLoS One* 9:e114211. doi: 10.1371/journal.pone.0114211

Veluchamy, A., Lin, X., Maumus, F., Rivarola, M., Bhavsar, J., Creasy, T., et al. (2013). Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat. Commun.* 4:2091. doi: 10.1038/ncomms3091

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., et al. (2015). An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. *Genom. Biol.* 16:102. doi: 10.1186/s13059-015-0671-8

Walker, G., Dorrell, R. G., Schlacht, A., and Dacks, J. B. (2011). Eukaryotic systematics: a user's guide for cell biologists and parasitologists. *Parasitology* 138, 1638–1663. doi: 10.1017/S0031182010001708

Wang, P., Gao, J., Wan, C., Zhang, F., Xu, Z., Huang, X., et al. (2010). Divinyl chlorophyll(ide) a can be converted to monovinyl chlorophyll(ide) a by a divinyl reductase in rice. *Plant Physiol.* 153, 994–1003. doi: 10.1104/pp.110.158477

Wang, Y., Jensen, L., Højrup, P., and Morse, D. (2005). Synthesis and degradation of dinoflagellate plastid-encoded psbA proteins are light-regulated, not circadian-regulated. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2844–2849. doi: 10.1073/pnas.0406522102

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2006). Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.* 34, 173–180. doi: 10.1093/nar/gkj158

Yu, M., Ashworth, M. P., Hajrah, N. H., Khiyami, M. A., Sabir, M. J., Alhebshi, A. M., et al. (2018). "Evolution of the plastid genomes in diatoms," in *Plastid Genome Evolution*, Vol. 83, eds S. M. Chaw and R. K. Jansen (Amsterdam: Elsevier), 129–155. doi: 10.1016/bs.abr.2017.11.009

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A., and Horvath, S. (2010). Weighted gene coexpression network analysis: state of the art. *J. Biopharm. Stat.* 20, 281–300. doi: 10.1080/10543400903572753

Zhao, X., Deton Cabanillas, A.-F., Veluchamy, A., Bowler, C., Vieira, F. R. J., and Tirichine, L. (2020). Probing the diversity of Polycomb and Trithorax proteins in cultured and environmentally sampled microalgae. *Front. Mar. Sci.* 7:3389. doi: 10.3389/fmars.2020.00189

Zheng, Y. T., Quinn, A. H., and Sriram, G. (2013). Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine diatom *Phaeodactylum tricornutum*. *Microb. Cell Fact.* 12:109. doi: 10.1186/1475-2859-12-109

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ait-Mohamed, Novák Vanclová, Joli, Liang, Zhao, Genovesio, Tirichine, Bowler and Dorrell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fpls.2020.578915

Comparative Proteomic Analysis Reveals New Insights Into the Common and Specific Metabolic Regulation of the Diatom *Skeletonema dohrnii* to the Silicate and Temperature Availability

Satheeswaran Thangaraj[1], *Mario Giordano*[2] *and Jun Sun*[1]*

[1] College of Marine Science and Technology, China University of Geosciences (Wuhan), Wuhan, China, [2] Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche, Ancona, Italy

OPEN ACCESS

Edited by:
Hanhua Hu,
Institute of Hydrobiology, Chinese Academy of Sciences, China

Reviewed by:
Weichao Huang,
Carnegie Institution for Science, United States
Yufang Pan,
Institute of Hydrobiology, Chinese Academy of Sciences, China

***Correspondence:**
Jun Sun
phytoplankton@163.com

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: 01 July 2020
Accepted: 28 September 2020
Published: 05 November 2020

Citation:
Thangaraj S, Giordano M and Sun J (2020) Comparative Proteomic Analysis Reveals New Insights Into the Common and Specific Metabolic Regulation of the Diatom Skeletonema dohrnii to the Silicate and Temperature Availability. Front. Plant Sci. 11:578915. doi: 10.3389/fpls.2020.578915

Silicate (Si) and temperature are essential drivers for diatom growth and development in the ocean. Response of diatoms to these particular stress has been investigated; however, their common and specific responses to regulate intracellular development and growth are not known. Here, we investigated the combination of physiological characteristics and comparative proteomics of the diatom *Skeletonema dohrnii* grown in silicate- and temperature-limited conditions. Results show that cell carbon and lipid quotas were higher at lower-temperature cells, whereas cellular phosphate was higher in cells grown with lower Si. In silicate-limited cells, nitrate transporters were downregulated and resulted in lower nitrate assimilation, whereas the phosphate transporters and its assimilation were reduced in lower-temperature conditions. In photosynthesis, lower silicate caused impact in the linear electron flow and NADPH production, whereas cycling electron transport and ATP production were affected by the lower temperature. Concerning cell cycle, imbalances in the translation process were observed in lower-silicate cells, whereas impact in the transcription mechanism was observed in lower-temperature cells. However, proteins associated with carbon fixation and photorespiration were downregulated in both stress conditions, while the carbohydrate and lipid synthesis proteins were upregulated. Our results showed new insights into the common and specific responses on the proteome and physiology of *S. dohrnii* to silicate and temperature limitation, providing particular nutrient (Si)- and temperature-dependent mechanisms in diatoms.

Keywords: silicate, temperature, diatom, iTRAQ, ribosome metabolism, photosynthesis, comparative proteomics, carbon metabolism

INTRODUCTION

Diatoms are key ecological players in the contemporary ocean, responsible for 40% of oceanic primary production (Falkowski et al., 1998); they contribute extensively to the global carbon cycle (Martin et al., 2011). They are possibly the most abundant silicified organisms on Earth (Shrestha et al., 2012) due to their silicon cell wall (Allen et al., 2011) believed to be the reason for their

ecological success (Hamm et al., 2003). Silicate is essential for diatoms not only because of their cell walls but also because of its requirement for diatom metabolism similar to other nutrients, i.e., nitrate, phosphate, and iron (Thangaraj et al., 2019). Temperature has a greater influence on algal metabolism (Bermudez et al., 2015), including diatoms (Depauw et al., 2012), because of its involvement in a high number of metabolic processes and enzymatic reactions. Recent reports demonstrated that changes in the temperature could alter the overall transcriptome (Liang et al., 2019) and proteome profile of diatom (Dong et al., 2016), leading to metabolic imbalances in amino acid biosynthesis and photosynthesis metabolism.

Silicon transporters (SITs) have recently been characterized in diatoms (Shrestha and Hildebrand, 2015); they are regulated by Si bioavailability, resulting in changes in gene expression and nutrient assimilation (Smith et al., 2016). Nutrient limitation could suppress the photosynthesis mechanism of phytoplankton depending on temperature (Maranon et al., 2018), showing that photosynthesis proteins and associated electron transport in diatom could change their nature with Si bioavailability (Thangaraj et al., 2019). In addition, transcriptome (Brembu et al., 2017) and proteome studies (Chen et al., 2018) on diatom's carbon metabolism have shown that changes in Si availability could decrease carbon fixation with simulation increase in the lipid accumulation for the acclimation process (Smith et al., 2016). The interrelation between amino acid biosynthesis and associated protein processing of ribosomal assembly in eukaryotes has been demonstrated (Lott et al., 2013) to alter the growth rate and cellular development (Zhou et al., 2015) under stressful conditions. Although cell cycle-related proteins were characterized in diatom response to Si deficiency (Du et al., 2014), it has not been explored whether the interconnection of amino acid biosynthesis and associated protein processing would alter growth and cellular development.

Temperature is essential for diatoms' photosynthetic process and enzymatic reactions (Huysman et al., 2013). In diatoms, each species has its own characteristic response to temperature fluctuation (Boyd and Hutchins, 2012) to evolve a sophisticated cellular mechanism for their ecological success (Brunet and Lavaud, 2010). Nevertheless, current understanding of the light-driven process in marine diatoms is still limited at the molecular level (Depauw et al., 2012), especially at lower temperatures (night) where algae consume 22% of their biomass (Edmundson and Huesemann, 2015) and exhibit NPQ as an important photo-protective process (Jallet et al., 2016) to manage their cellular energy and dissipation. The transcriptome response of diatoms has shown that, under lower temperatures, amino acid biosynthesis and ribosome complex altered their functions (Liang et al., 2019). In contrast, higher temperatures stimulated changes in the diatom's photosynthesis and associated electron transport along with the light-harvesting complex (Dong et al., 2016). Maranon et al. (2018) reported that nutrient uptake at the molecular level can be influenced by temperature limitation, although it is the consequence of the interplay of nutrient availability and transporters. The transcriptome (Liang et al., 2019; Thangaraj and Sun, 2020a) and proteomic changes (Dong et al., 2016) of a diatom's response to changing temperature showed that temperature is the dominant factor altering gene expression in cell cycle progress and protein synthesis for cellular development and growth. Also, metabolic changes in multiple diatoms revealed that changes in carbon metabolism could be temperature-dependent, determining decreased photosynthesis and carbon fixation with increased lipid accumulation for their acclimation process (Liang et al., 2019).

The ecological success of diatoms has determined the emergence of an interest in the understanding of the regulation of their proteome in response to a variety of environmental parameters: nitrate (Jian et al., 2017), phosphate (Dyhrman et al., 2012), silicate (Du et al., 2014), iron (Nunn et al., 2013), and availability of temperature (Dong et al., 2016). Although the outcome of these studies has shown significant metabolic regulation or proteome exposed to specific factors, there is a lack of information on the interplay of different environmental effectors and their cumulative effect on cellular growth and development. In order to address this matter, iTRAQ-based comparative proteomics was applied in this study to compare the proteome profile of the diatom *Skeletonema dohrnii* grown in silicate and temperature limitation. Temperature was chosen because of it influenced metabolism in a non-specific mode (although it may elicit specific responses for the amelioration of stresses linked to certain functions), whereas silicon is specifically related the construction and maintenance of the involucral system of the cell.

MATERIALS AND METHODS

Sampling and Culture Condition

This study was conducted using ecologically important diatom *S. dohrnii*, isolated from the Yellow Sea. The isolated strain was maintained in Artificial Seawater Media (ASW) (Sunda et al., 2005), at 25°C with an irradiance of 100–120 μmol photons m^{-2} s^{-1}, at a 12:12-h light–dark cycle (Thangaraj and Sun, 2020b). For the experiment, cultures were grown in two different temperatures (control 25°C and lower temperature 15°C) and silicate concentrations (control 1 ml/L and lower silicate 0.2 ml/L), respectively, with the abovementioned conditions. The final concertation of used silicate stock solution in this study was 229.71 μM/ml. To prevent the extra silicon utilization of *S. dohrnii* during the experiment, cells were grown in Nalgene, polycarbonate bottles for the silicate treatment. Both silicate- and temperature-limited cells were harvested during mid-exponential growth (day 4) filtering through 2-mm pore-size membrane filter for the cellular elements (C, N, P, and Si), macromolecular components (carbohydrate, lipids, and protein), and quantitative proteomic analysis.

Measurements of Growth Rate, Nutrient Analysis, and Cell Constituents

The reproducibility of each condition was checked by using three independent triplicate samples, for cell growth, nutrients analysis,

cellular elements, and biochemical properties. The growth rate μ (day^{-1}) in this study was calculated as follows:

$$\mu = 1n(N_t : N_0)/t$$

where, N_0 and N_t are at the end and start of the exponential phase of growth, respectively, and t is the duration of the exponential growth phase. The concentration of nutrients (N, P, and Si) in each medium of culture was measured following continuous flow analysis (CFASAN Plus/Skalar Analytik, Germany). To determine the cellular elements (C, N, P, and Si), culture aliquots containing about 2×10^7 cells were collected and dried in an oven at 65°C for 2 days, prior to pulverization in a tissue layer (Elementar, Germany). For N, P, and Si, the pulverized samples were transferred into a tinfoil cup and analyzed with a photometric auto-analyzer (CFASAN Plus/Skalar Analytic, Germany) and EL cube (Elementar, Germany) following the protocol described by Boyd et al. (2015); Chen et al. (2018). For cellular elements, the total carbohydrate was analyzed using the anthrone method (Guerra et al., 2013); total lipids were measured according to protocol described by Yoneda et al. (2018) and determined gravimetrically using a microanalytical balance. The total proteins were determined following the Folin-Phenol method of Lowry et al. (1951). Chlorophyll *a* was extracted in 90% acetone, at 4°C for 24 h in the dark and quantified spectrophotometrically using the equation in Jeffrey and Humphrey (1975). All measurements were carried out on three biological replicates; statistical significance was assessed by *t*-test.

Protein Extraction, Preparation, and Digestion

One liter of culture from each sample was collected through 2 μM filter membrane; then, subsequently, samples were suspended in 10 ml of medium into 15-ml centrifuge tubes for protein preparation (Du et al., 2014). The resulting cell pellets were then suspended in lysis buffer 3 (8 M Urea, 40 mM Tris-HCl or TEAB with 1 mM PMSF, 2 mM EDTA, and 10 mM DTT, pH 8.5). The mixture of samples was placed into a tissue lyser for 2 min at 50 Hz to cell lysis and then centrifuged at 25,000*g* for 20 min at 4°C. The supernatant was then transferred into a new tube; samples were reduced with 10 mM dithiothreitol (DTT) at 56°C for 1 h and alkylated by 55 mM iodoacetamide (IAM) in the dark at room temperature for 45 min to block the cysteine residues of the proteins. Following centrifugation (25,000*g* for 20 min at 4°C), the supernatant containing proteins was quantified by the Bradford assay method (Kruger, 2009). The protein solution (100 μg) with 8 M urea was diluted four times with 100 mM TEAB. Trypsin Gold (Promega, Madison, WI, United States) was used for the protein digestion with a ratio of trypsin = 40:1 at 37°C overnight. After trypsin digestion, peptides were desalted with a Strata X C18 column (Phenomenex) and vacuum-dried according to the manufacturer's protocol for 8-plex iTRAQ (Applied Biosystems, Foster City, CA, United States). Each treatment was made up of two (control) or three (treated with lower Si or lower temperature) biological replicates. Briefly, peptides were labeled with iTRAQ reagents 113 and 115 for control samples; 114, 116, and 118 for lower-silicate samples; and 117, 119, and 121 for lower-temperature samples, and then pooled and dried by vacuum centrifugation.

Analytical Procedure and Peptide Labeling

The labeled peptide blends were pooled and dried through vacuum centrifugation and fractionated. All solvents used for high-performance liquid chromatography (HPLC) were HPLC grade (Sigma-Aldrich), and the H_2O was Millipore Milli-Q PF filtered. The peptides were separated on a Shimadzu LC-20AB HPLC Pump system coupled with a high-pH reverse-phase column (Gemini C_{18} 5 μM, 4.6 × 250 mm). The peptides were reconstituted to HPLC separation with the following mobile phase: (A) 5% ACN, (B) 95% H_2O (adjusted pH to 9.8 with 2 ml of NH_3), sample input and acquisition: 2 ml/min flow rate and 1 ml/min injection volume. Crude peptide compound elution was monitored by measuring UV absorbance at 214 nm, and the 40 fractions were collected every 1 min. All the eluted peptides were combined as 20 fractions and vacuum-dried for further process. Furthermore, each fraction was resuspended in buffer A (2% ACN and 0.1% formic acid in H_2O) and then centrifuged at 20,000*g* for 10 min and independently subjected to HPLC separation (LC-20AD nano-HPLC instrument, Shimadzu, Kyoto, Japan) using C_{18} column (inner diameter, 75 μm). Sample input and acquisition: 300 nl/min flow rate and 1 μl injection volume for 8 min; the 35-min gradient was run at 300 nl/min starting from 8 to 35% of buffer B (2% H_2O and 0.1% FA in ACN), followed by a 5-min linear gradient to 80% solution B, maintenance at 80% solution B for 4 min, and return to 5% in 0.1 min and equilibrated for 10 min.

LC-MS/MS Proteomic Analysis

Liquid chromatography and mass spectrometry (LC-MS) analysis of diatom peptide was performed on LC-20AD (Shimadzu, Kyoto, Japan) using C_{18} column (size, 75 μm). The LC-MS data were acquired in positive ion mode with a selected mass range of 350–1500 *m/z*. Based on the intensity in the MS1 survey, as many as 30 production scans were collected if beyond a threshold of 120 counts per second (counts/s) and charge state 2+ to 5+ dynamic exclusion was set for 1/2 of peak width (12 s). For MS data acquisition, the collision energy was adjusted to all precursor ions for collision-induced dissociation, and the Q2 transmission window for 100 Da was 100%. The greatest extents of the iTRAQ reporter ions imitate the relative abundance of the proteins in the samples. A TripleTOF 5600 mass spectrometer with high mass accuracy resolution (less than 2 ppm) was used in this study for peptide identification. Other identification parameters used included the following: fragment mass tolerance: ± 0.1 Da; mass values: monoisotopic; variable modifications: Gln- > pyro-Glu (N-term Q), oxidation (M), iTRAQ8plex (Y); peptide mass tolerance: 0.05 Da; max missed cleavages: 1; fixed modifications:

carbamidomethyl (C), iTRAQ8plex (N-term), iTRAQ8plex (K); other parameters: default.

Proteomic Data Analysis

All the mass spectral data were processed using the Proteo Wizard software msConvert with default parameters for generating peak list, and the data alignment was performed with Analyst QS 2.0 software (Applied Biosystems/MDS SCIEX). Further, protein identification and quantification were achieved using Mascot 2.3.02 (Matrix Science, London, United Kingdom) (Charbonneau et al., 2007). For iTRAQ quantification, the peptide for quantification was automatically selected by the algorithm to calculate the reporter peak area (using default parameters in the Mascot Software package). The acquired data were auto bias-corrected to get rid of any differences imparted due to the unequal mixing during combining differently labeled samples. Proteins with a 1.2-fold change between each different sample and a p value of statistical evaluation less than 0.05 were determined as differentially expressed proteins (DEPs). The Student's t-test was performed using the mascot 2.3.02 software. Briefly, a protein ratio is reported in boldface if it is significantly different from unity. The comparison test is:

$$|X - \_| \_t * \frac{S}{\sqrt{N}}$$

If this dissimilarity is real, then there is no important difference at the stated sureness level. Further, N is the number of peptide ratios, S is the standard deviation, and X is the mean of the peptide ratios, with both numbers calculated in log space. The real value of the ratio, μ, is 0 in log space. t is Student's t for N - 1 degrees of freedom and a two-sided confidence level of 95%.

Functional Annotation

The COG (Cluster of Orthologous Groups of proteins) and then GO (Gene Ontology) analyses were performed according to the method reported by Unwin (2010). The identification of differentially regulated proteins in GO terms was carried out using the following formula:

$$P = 1 - \hat{a}_{i=0}^{m-1} \frac{\left(\frac{M}{i}\right)\left(\frac{N-M}{n-i}\right)}{\left(\frac{N}{n}\right)}$$

Where N is the number of all proteins with GO annotation information, n is the number of the differentially regulated proteins with GO annotation information, M is the number of proteins with a given GO term annotation, and m is the number of the differentially regulated proteins with a given GO term annotation. The GO terms with a p value of less than 0.05 were considered as enriched GO terms by the silicate- and temperature-responsive proteins. Proteins with twofold changes between each sample and a p value of less than 0.05 were considered as differentially expressed. The metabolic pathway analysis of DEPs was conducted according to the KEGG Pathway Database (Kanehisa et al., 2016).

RESULTS

Physiological and Biochemical Responses

Both lower Si and temperature resulted in significant changes in the growth rate of the diatom *S. dohrnii* (**Figure 1A**), showing 50% reduction in cells grown with lower Si (0.8 ± 0.1) and 45% reduction at lower temperatures (0.9 ± 0.2) compared with the normal condition (1.6 ± 0.2). The concentration of nutrients (N, P, and Si) in the culture media of all the conditions is shown in **Figures 1B–D**. Results show that compared with the normal culture condition, the concentration of N and Si is lower in Si-limited cells, whereas P was lower in cells limited by temperature.

The cell quotas of the main macronutrients and organic pools are shown in **Figure 2**, for cells grown at lower-silicate, lower-temperature, and control conditions. In the Si treatment, C quota was 20% lower than in the control cells; the N and Si cell quotas were much lower than in the controls, whereas P content was higher. In lower-temperature treatment, C quota was 12% higher than in the Si treatment and N was lower than in the control but was double as compared to Si-limited cells. The P cell quota was much lower in the temperature treatment cells than in both control and Si treatment, and Si cell quotas were the lowest in temperature treatment with respect to organic pools. Compared with control cells, both silicate and temperature treatment cells have increased the size of carbohydrates and lipid pools but decreased protein synthesis.

Overview of Quantitative Proteomics

In total, 343,191 spectra were collected, corresponding to 3479 peptides and 1772 proteins that were identified with 1% of FDR. Detailed information—accession numbers, protein descriptions, unique peptide, spectrum, p-Value, and q-Value with repeatability analysis between the samples—is given in **Supplementary Table 1**. Protein abundances changed by at least twofold with p values lower than 0.05; 1380 proteins were differentially expressed in this study between silicate and temperature limitation (**Supplementary Table 2**). In lower-silicate cells, 411 proteins were downregulated and 316 were upregulated, and in the lower-temperature cells, 291 proteins were upregulated and 362 were downregulated. Among them, 359 proteins were commonly expressed in both conditions, of which 199 were downregulated and 160 were upregulated (**Figure 3A**). These DEPs were then plotted using the volcano plot analysis, shown in **Figures 3B,C**.

Functional Classification of COG and GO and KEGG Pathway Enrichment Analysis

In total, 12,378 unigenes were identified in the COG database; they were attributable to 22 categories, based on sequence homology (**Figure 4A**). The largest group included proteins involved in ribosomes and their biogenesis (18%), followed by protein involved in post-transitional modification, protein turnover and chaperones (13%), and amino acid transport and metabolism (11%). To further understand the specific functions of these unigenes, they were subject to GO analysis;

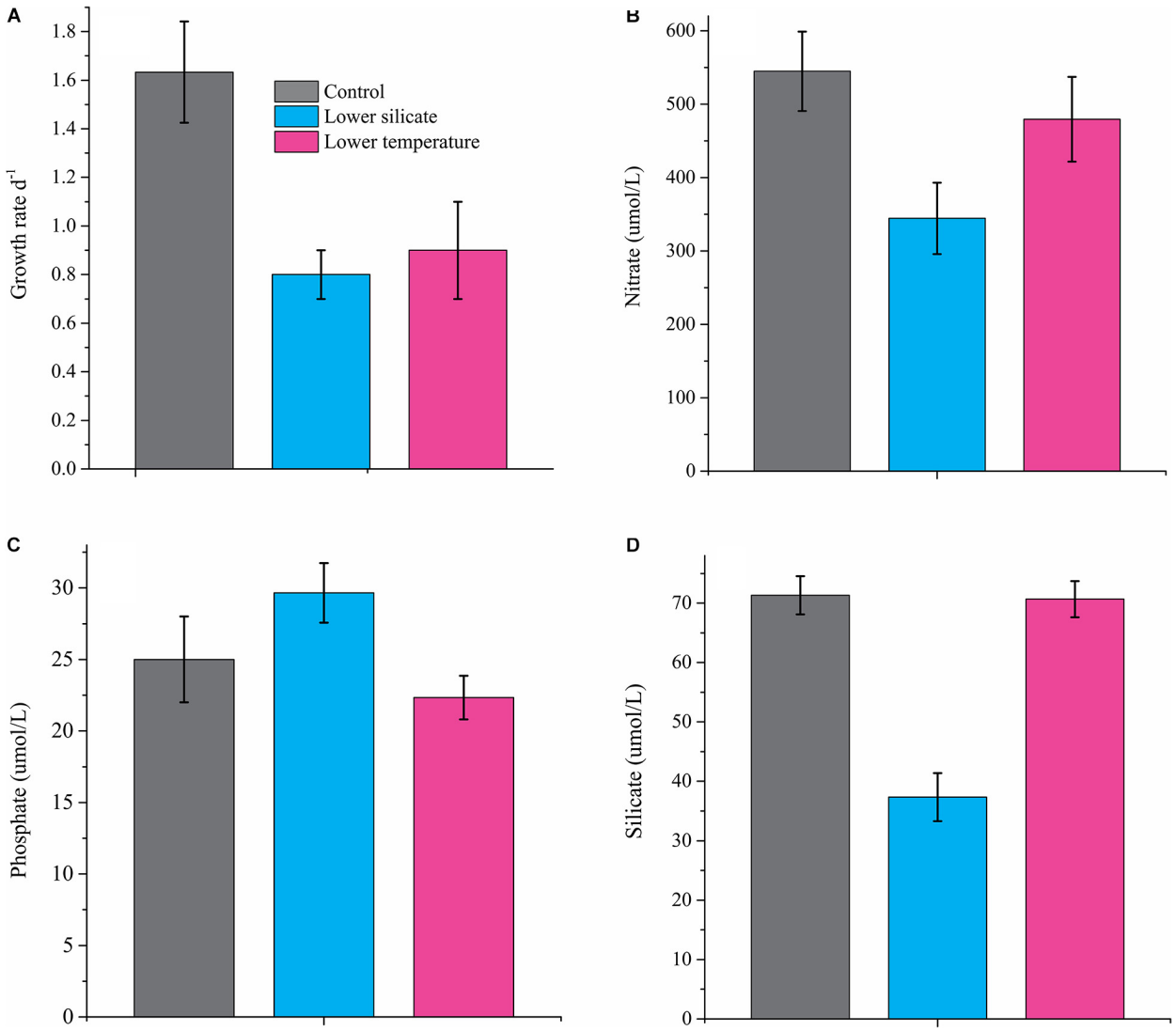


FIGURE 1 | The specific growth rate of *S. dohrnii* grown under normal conditions, silicate and temperature limitation **(A)**, and external nutrient concentration, nitrate **(B)**, phosphate **(C)**, and silicate **(D)** in culture media. Error bars represent the standard deviation of the means (n = 3).

they were classified into three ontologies and 37 sub-categories (**Figure 4B**). The most represented ontology in molecular function was a catalytic activity, associated with photosynthetic electron transport. In terms of structural components, proteins involved in the cell cyclic, photosynthetic membranes and chloroplast thylakoid membrane are the most represented. For the biological process, a large portion of the protein identified is related to metabolic regulations of ATP synthesis, proton transport, and glycolysis. To understand the DEP involvement into the specific biological pathway, KEGG pathway enrichment analysis was carried out and shows (**Figure 5**) the most represented pathways in both silicate- and temperature-limited cells. Pathways of photosynthesis, light-harvesting, oxidative phosphorylation, carbon fixation, carbon metabolism, amino acid biosynthesis, and ribosome metabolism were involved in most DEPs in both lower-silicate and lower-temperature cells, although the number of proteins and fold change in each pathway is different.

Changes in the Nutrients Transport

In lower-silicate cells, two silicon transporters (SIT1 and SIT2) were upregulated and 10 nitrate transporters, including Nitrilase (NIT2), NRT2, Urease (URE), and glutamate dehydrogenase, were downregulated (**Table 1**). In cells grown at lower temperatures, three transporters associated with nitrate assimilation, i.e., nitrate transporter (NRT1), urea protein transporter (DUR3), and xanthine uracil permease (TN.NCS2), and phosphate transporters [vacuolar transporter chaperone 4 (VTC4), and 5′-nucleotidase (ushA)] were downregulated. In addition, fewer sulfate transporters, methionine S-adenosyl transferase, and sulfate transporters were also noticed being downregulated in cells treated at lower temperatures; however,

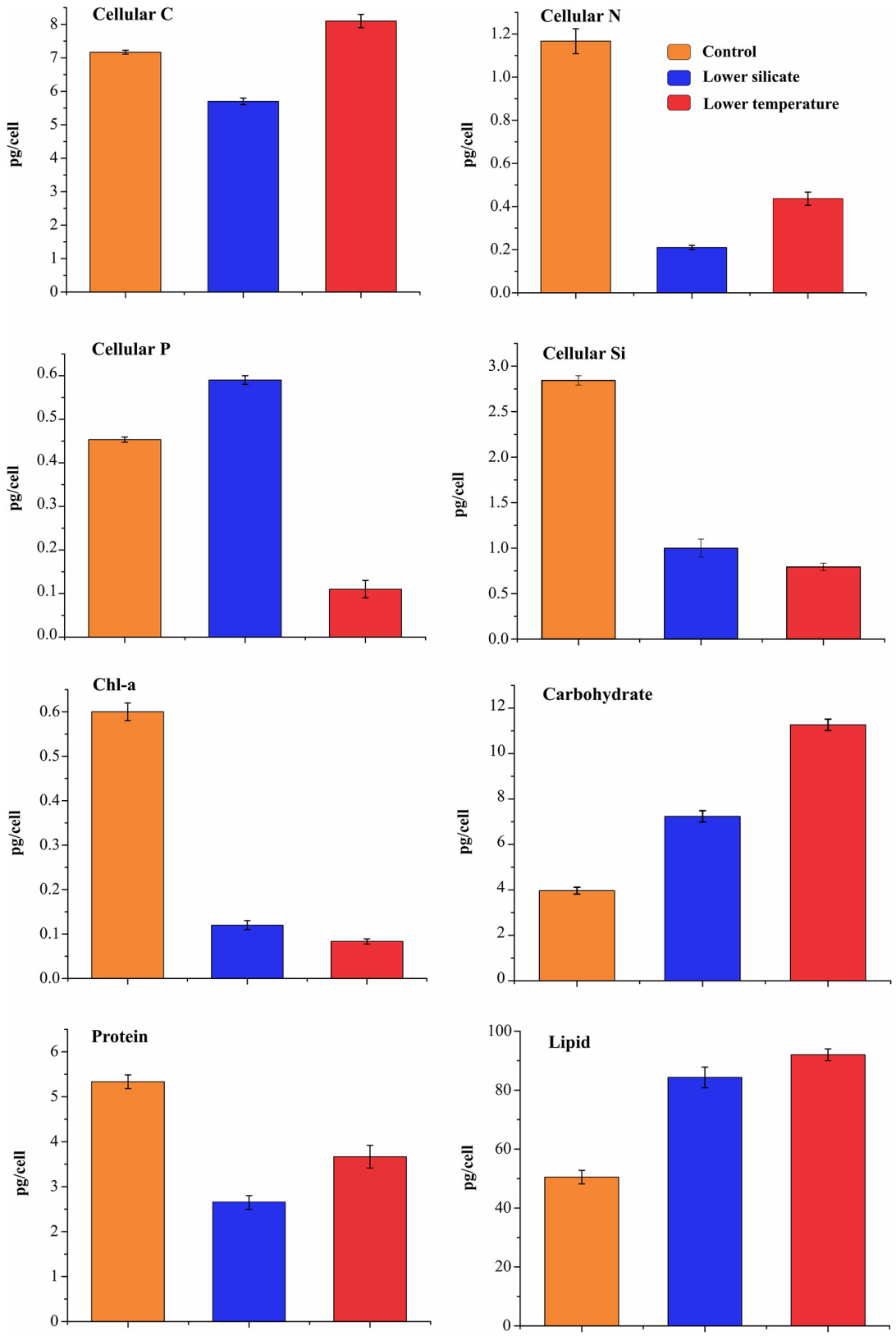


FIGURE 2 | Cellular elements and macromolecular component composition in *S. dohrnii* grown under control, silicate-, and temperature-limited conditions. The error bars represent the standard deviation of the means (n = 3).

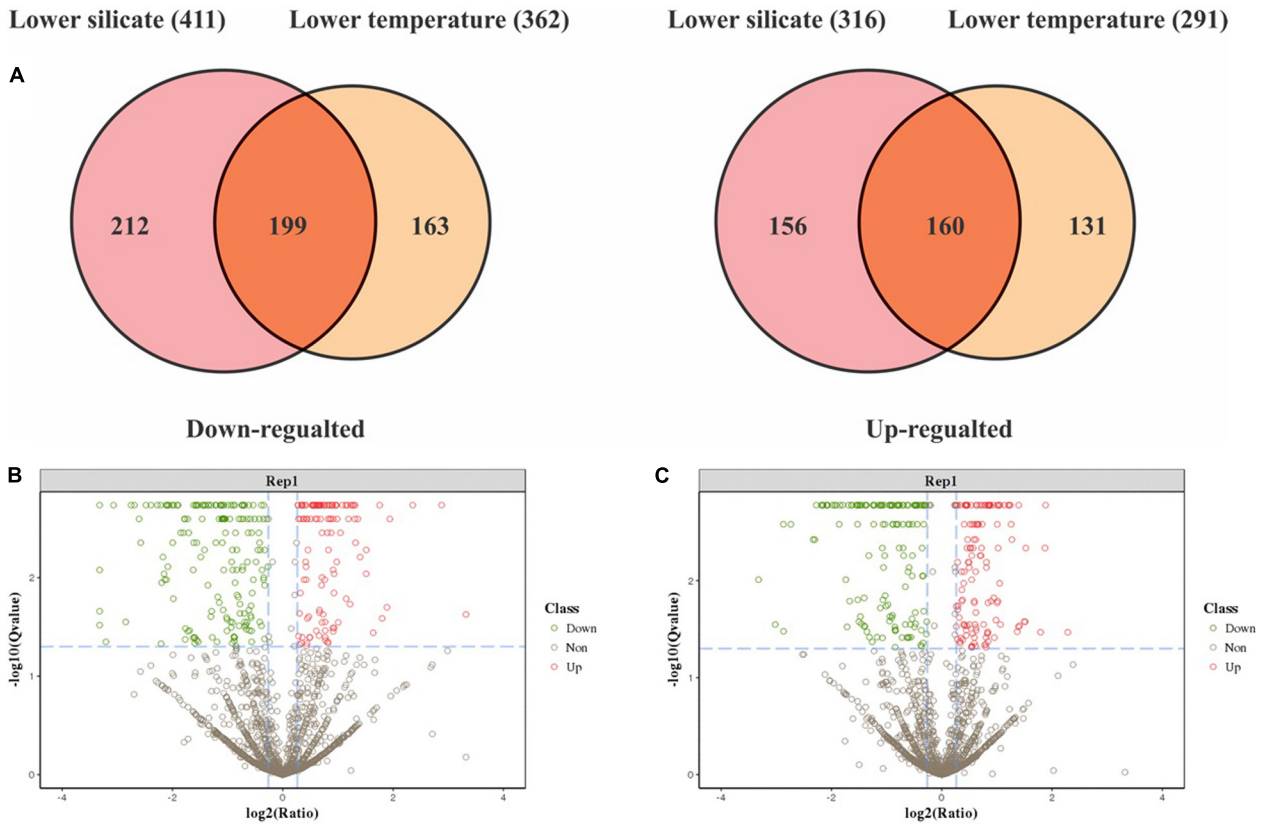


FIGURE 3 | The Venn diagram depicts common and specific differentially expressed proteomic responses between silicate and temperature limitation **(A)**, and volcano plots of differentially expressed proteins in lower-silicate cells **(B)**, and lower-temperature cells **(C)**.

this has not been differentially expressed in cells treated with lower silicate.

Changes in the Photosynthesis Metabolism

In this study, 20 downregulated proteins associated with photosynthesis were identified in the Si-limited treatment and 23 were identified in the lower-temperature treatment (**Supplementary Table 3**). Of these downregulated proteins, six and eight proteins specifically responded to silicate and temperature limitations, whereas 14 proteins responded commonly at both stress conditions. Downregulated proteins were classified into five groups: Photosystem II (PSII), Photosystem I (PSI), Light-harvesting complexes (LHC), Photosynthetic electron transport, and Chloroplast F-type ATPase. Quantitative proteomics revealed that six PSII proteins (*Psb A, B, C, D, E, H*), two PSI proteins (*Pas A, B*), six LHC proteins, and four F-type ATPase proteins were downregulated in lower Si-grown cells, whereas in lower-temperature cells, four PSII proteins (*Psb B, E, D, H*), two cytochrome proteins, one protein in PSI, and 10 LHCs were downregulated. Though many of the proteins were affected by both treatments, the extent of the regulation (fold change) varied: lower silicate appears to have a stronger impact on the PSII complex proteins, while the lower temperature had more influence on the LHC proteins.

Changes in the Carbon Fixation and Carbohydrate Metabolism

In total, 41 proteins involved in carbon metabolism were differentially expressed in the lower Si treatment; in cells acclimated to lower temperature, 42 proteins showed changes in abundance relative to control (**Supplementary Table 3**). Of this DEP, 22 proteins were specifically expressed in lower-silicate cells, whereas 21 proteins were expressed in cells treated with lower temperature; 20 common DEPs were identified in both stress conditions. In cells treated with lower silicate, an important enzyme of glyceraldehyde-3-phosphate dehydrogenase (GAPD) was downregulated, while this protein was upregulated in cells treated at lower temperatures. Further, fumarate and pyruvate kinase (PYK) were upregulated in lower Si-grown cells, while these proteins in temperature-limited cells did not express differentially. In lower-temperature cells, phosphofructokinase (pfkA) was also upregulated, and so were some proteins related to pentose phosphate pathway (xylulose, sedoheptulose, and erythrose), but these genes were not expressed differentially in cells exposed to lower silicate. In addition, the specific response of lower temperature caused downregulation of oxaloacetate

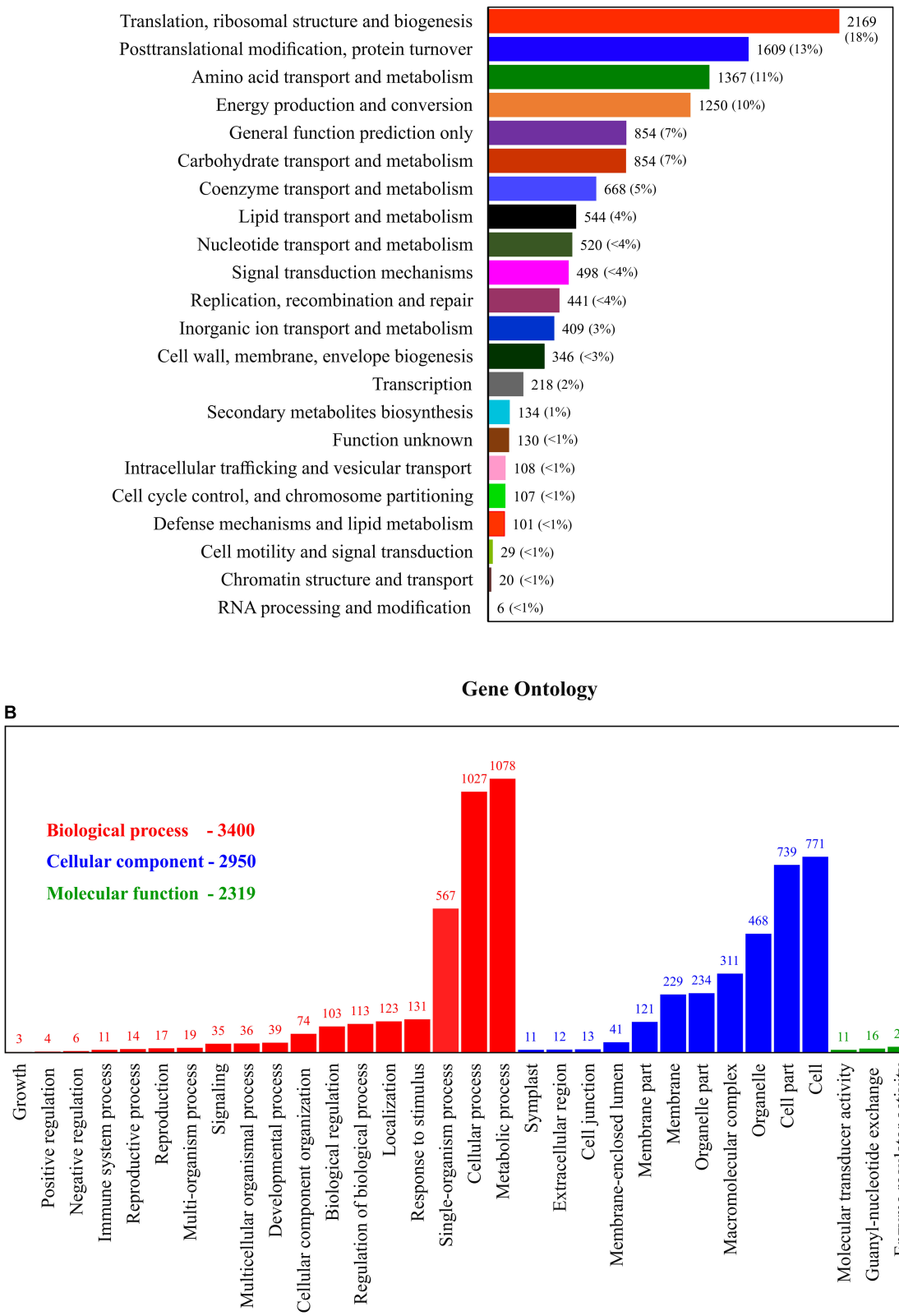


FIGURE 4 | Functional annotation of the COG **(A)** and GO **(B)** analysis of the proteins identified in this study.

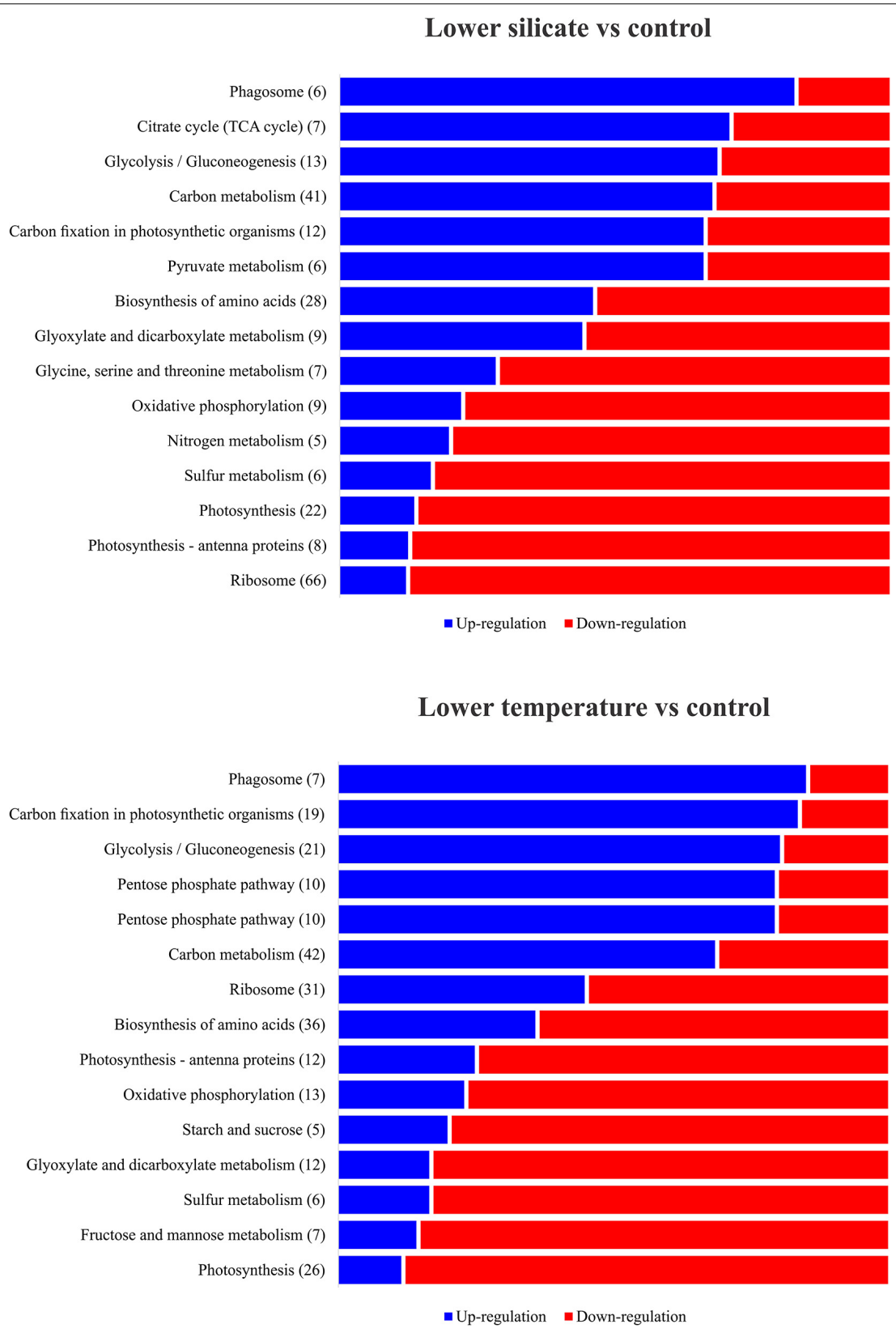


FIGURE 5 | Most represented KEGG pathway enrichment analysis of differentially expressed proteins in the diatom *S. dohrnii*; those significantly enriched ($p < 0.05$) under lower-silicate and lower-temperature cells were denoted as upregulated (blue) and downregulated (red) within each pathway.

TABLE 1 | Differentially abundant proteins/transporters associated with the utilization of silicon, nitrate, phosphate, and sulfate in *S. dohrnii* treated with lower silicate and temperature.

Transporter Name	Accession ID	Unique peptides	Fold change
Lower silicate vs. control			
Silicate utilization			
Silicic acid transporter (SIT1)	XP_002290700.1	3	4.23
Silicic acid transporter (SIT2)	XP_002295920.1	5	3.41
Nitrate utilization			
Nitrate/nitrite transporter (NRT1)	AAV67002.1	6	−1.26
Nitrate/nitrite transporter (NRT2)	EJK46860.1	3	−1.67
Glutamate dehydrogenase	XP_002289225.1	2	−0.82
Aliphatic amidase	XP_002289996.1	3	−1.22
Nitrilase (NIT2)	XP_002290043.1	4	−0.91
AMP deaminase	XP_002289781.1	6	−1.41
5-hydroxyisourate hydrolase	XP_002288652.1	11	−1.19
Allantoicase	XP_002289615.1	2	−1.81
NADPH Nitrite reductase	XP_002287665.1	13	−1.14
Urease (URE)	XP_002296690.1	9	−1.23
Lower temperature vs. control			
Nitrate utilization			
Nitrate/nitrite transporter (NRT1)	EED92802.1	4	−0.68
3 Urea-proton symporters (DUR3)	XP_002292926.1	4	−1.71
Xanthine/uracil permease (TN. NCS2)	EED96094.1	7	−1.14
Phosphate utilization			
Vacuolar transporter chaperone 4 (VTC4)	EED87388.1	5	−1.43
Phospholipase D1/2 (PLD1_2)	XP_002288407.1	3	−2.17
5'-nucleotidase	XP_002295180.1	2	−1.87
Betaine aldehyde dehydrogenase	ACI64514.1	11	−1.31
Sulfur utilization			
Phosphoadenosine-phosphosulfate reductase	EED88796.1	6	−1.39
Methionine S-adenosyl transferase	BAH30220.1	6	−2.24
Sulfate transporter	XP_002286457.1	2	−1.84

Detailed annotation of these nutrient transporters is given in ***Supplementary Table 2****.*

(OAA), and upregulation of sulfotransferase (ST), whereas these proteins (OAA) and ST abundances were up- and downregulated, respectively, in lower-silicate cells.

Some of the proteins involved in CO_2 fixation were downregulated at both stress conditions, such as carbonic anhydrase (CA), ribulose-1,5-bisphosphate carboxylase (rbcL), phosphoenolpyruvate carboxylase (PEPC), and pyruvate phosphate dikinase (PPDK). Similarly, transketolase (TST), methionine chain elongation, (BCAT4), and S-adenosylmethionine synthetase (SAM) were downregulated at both stress conditions. Notably, key genes involved in lipid biosynthesis, i.e., acetyl-CoA carboxylase (ACACA), and long-chain fatty acyl-CoA were upregulated along with isocitrate (ICDH), pyruvate dehydrogenase (PYD), and succinate (SDH). Moreover, many proteins involved in glycolysis, TCA, and amino acid biosynthesis—phosphoglycerate mutase (PGM), glucose-6-phosphate isomerase (GPI), phosphoglycerate kinase (PGK), pyruvate dehydrogenase (PDC), fructose-bisphosphate aldolase (FBA), and citrate synthase (CS)—were similarly expressed in both stress conditions, although to a different extent.

Changes in the Ribosome Metabolism

In total, 66 ribosomal proteins consisting of 47 large subunit and 19 small subunit proteins were differentially expressed in cells cultured with lower silicate (**Supplementary Table 3**). Among them, 45 proteins in large subunits and 16 proteins in small subunits were downregulated. Of these downregulated proteins, 10 were proteins associated with chloroplast in the large subunit and 4 were proteins related to 40S and 60S in the small subunit. Similarly, in lower-temperature cells, 31 proteins were differentially expressed, among them 15 proteins in large subunits and 3 proteins in small subunits were downregulated. Notably, 47 proteins were specifically expressed in lower-silicate cells, 19 proteins were expressed in cells treated at lower temperatures, whereas 22 proteins were commonly expressed in both stress conditions, although to a different extent.

Changes in the Cell Cycle and Nucleus Related Proteins

In this study, many proteins related to cell cycle and nucleus were identified, and most of them were being downregulated. In lower Si treatment, a total of 49 proteins related (r, m, and t-RNA) to the ribosome, transporting coding sequences and translating information to protein, were downregulated (**Supplementary Table 2**). Cells exposed to lower temperatures downregulated 29 proteins related to DNA binding and its transcription to RNA (**Supplementary Table 2**). Specifically, high light-induced proteins and cullin protein were downregulated in cells acclimated to lower temperature. Similarly, dead box RNA helicase and GTP proteins were downregulated in the Si-deficient cells. An important cell cycle control protein, casein kinase II, was downregulated in low Si, while it was not expressed in lower-temperature cells. In lower-temperature cells, a core histone (H3) and one variant (H2A) associated with DNA binding and histone arginine N-methyltransferase were downregulated; in the Si treatment, histones (H3, H4) and histone deacetylase were downregulated. Proliferating cell nuclear antigen (PCNA), cyclin-dependent kinase (CDK), and mismatch repair (MMR) protein were decreased in both Si- and temperature-limited cells with varied downregulation ratios.

DISCUSSION

Specific Cellular Nutrient Assimilation

In general, cellular functional efficiency relies on nutrient transporters, which define what metabolites and compounds can across the membrane. In this study, changes in the Si concentration and temperature availability can regulated not only nutrient transporters (**Table 1**) but also the utilization of organic nutrients (**Figure 1**). Recent transcriptome and proteomic studies on diatoms have shown various SIT upregulation during the silicon limitation (Shrestha et al., 2012; Du et al., 2014;

Smith et al., 2016; Brembu et al., 2017). The outcome of these studies has shown that during silicon limitation, SITs act as a transcriptional cascade with encoding unknown amino acids to facilitate the silicic acid transport. Simultaneously, these encoded SITs have a unique subfunction in diatom metabolism, preventing translation into full-length protein and cell cycle arrest (Shrestha and Hildebrand, 2015; Durkin et al., 2016). These investigations have concluded that during silicon limitation, SITs have interrelation with cytoplasm to facilitate silicic acid transport, with the simultaneous improper binding of other functional proteins.

Nitrate is often mentioned to limit diatom growth in the ocean because of its contribution to binding chlorophyll, amino acids, and nucleic acids. Earlier proteomic observation on diatoms showed that limitation of iron, phosphate, and silicate has a coupling effect with N transporters, leading to downregulated nitrate transporters and their assimilation (Nunn et al., 2013; Chen et al., 2018). Similarly, many downregulations of nitrate transporters observed in this study responded to Si deficiency (**Table 1**), indicating lower nitrate assimilation (**Figure 1**) and cellular N quota (**Figure 1**). In any form of cell, extracellular N must convert into ammonium before assimilation into amino acids or other nitrogenous compounds. In this study, decreased abundance of ammonium converted transporters URE and AMP deaminase in Si-limited cells, suggesting a possible reduction of ammonium conversation and therefore N utilization. Proteins of Nitrilase (NIT2) and Aliphatic amidase play an essential role in N utilization (Howden and Preston, 2009), and downregulation of these transporters in Si-deficient cells also supports the reduction of N assimilation. In addition, allantoicase and 5-hydroxyisourate hydrolase involved in the purine utilization were also being downregulated in Si-limited cells, which is a substitute of N source for diatoms during stress conditions, and could regulate cellular development (Alipanah et al., 2015). Besides, decreased abundances of glutamate dehydrogenase under Si-limited conditions also led to decreased glutamate synthase, which has a critical role in intracellular N flow, as both N accepter and donor (Miflin and Habash, 2002). It is clear that *S. dohrnii* responded to lower Si caused and limited the N assimilation and utilization for the associated cellular process.

Similar to nitrate, phosphate also essential for amino acid and nucleic acid binding; in the ocean, polyphosphate is thought to be the product of phosphate storage in diatoms (Martin et al., 2011; Dyhrman et al., 2012). Accordingly, proteins containing polyphosphate synthase subunit (VTC4) are essential for maintaining intracellular phosphate homeostasis (Secco et al., 2012). In this study, cells treated at lower temperatures caused downregulation of VTC4 protein, suggesting that *S. dohrnii* was unable to store phosphate to fulfill its intracellular requirement. This is consistent with transcriptome (Zhang et al., 2016) and proteome response (Dyhrman et al., 2012) of diatoms *Skeletonema costatum* and *Thalassiosira pseudonana* to changing phosphate conditions. Diatoms can replace phospholipid with sulfur-containing (sulfolipids) and nitrogen-containing (betaine lipids) ones during phosphate deficiency to reduce the intracellular phosphate requirement of cells (Van Mooy et al., 2009). Downregulation of these transporters in lower-temperature cells suggests that *S. dohrnii* was unable to utilize non-phosphate-containing lipids and therefore affects the phosphate-associated cellular process. This finding is consistent with the earlier response of diatoms *S. costatum* (Zhang et al., 2016), *T. pseudonana* (Dyhrman et al., 2012), and *Chaetoceros affins* (Van Mooy et al., 2009) to changing phosphate conditions. The results on intracellular phosphate content (**Figure 2**) and nutrient concentration (**Figure 1**) in lower-temperature cells supported this speculation. In addition, the lower temperature also caused downregulation of three sulfur transporters, indicating the inadequate synthesized amount of cysteine, methionine, and glutathione with ultimate impact of growth rate and cellular development of diatom (Takahashi et al., 2011).

Specific Cellular Response in Photosynthesis and Associated Metabolism

Nitrate and phosphate are important for many physiological processes, including essential elements for protein synthesis. Reduction of these two nutrients' assimilation could cause a decreased synthesis rate of photosynthetic proteins (Bucciarelli and Sunda, 2003). Accordingly, in cells treated with lower silicate, PsbA, PsbB, PsbC, and PsbD proteins were downregulated, with those being assembled in photosystem II acting as a reaction center with complex light-harvesting also involved in the production and transfer of linear electron flow for the NADPH production (**Figure 6**). In cells grown under Si-deficient conditions, these proteins were downregulated $>$ 2-fold, whereas in lower-temperature treatment, it decreased $<$ 1-fold (**Supplementary Table 3**), indicating higher influences on the PSII complex and associated changes in the linear electron flow by Si limitation than temperature. Similar findings were seen in the proteome level on diatoms responding to iron (Nunn et al., 2013), nitrate (Hockin et al., 2012), and silicate limitation (Thangaraj et al., 2019), showing that downregulation of PSII proteins leads to impact the linear electron transfer and consequent oxidation of PSI complex and NADPH production.

Despite chlorophyll and fucoxanthin being rich with nitrogenous compounds, lower temperature regulated higher LHC and PSI proteins with greater decreases in protein abundances compared to Si limitation (**Supplementary Table 3**), suggesting that light-capturing capabilities were compromised with lower temperature. PSI and LHC proteins are involved in binding chlorophyll and catalyze light-induced photochemical processes. For example, PsbF is involved in increasing electron transfers and iron-binding processes, while fucoxanthin and chlorophyll proteins encompass the light-harvesting processes. In this study, decreased abundance of these proteins in lower-temperature cells may cause regulation in the cyclic electron flow (**Figure 6**), leading to redox imbalance in photosynthesis metabolism. Non-photochemical quenching (NPQ) and alternative electron transport (AEF) have been carried out by algae to overcome the redox imbalance. NPQ is an important photo-protective process that dissipates excess energy during light changing conditions (Niyogi, 2000). To avoid energetically costly damage to the cell, diatoms have evolved a series of

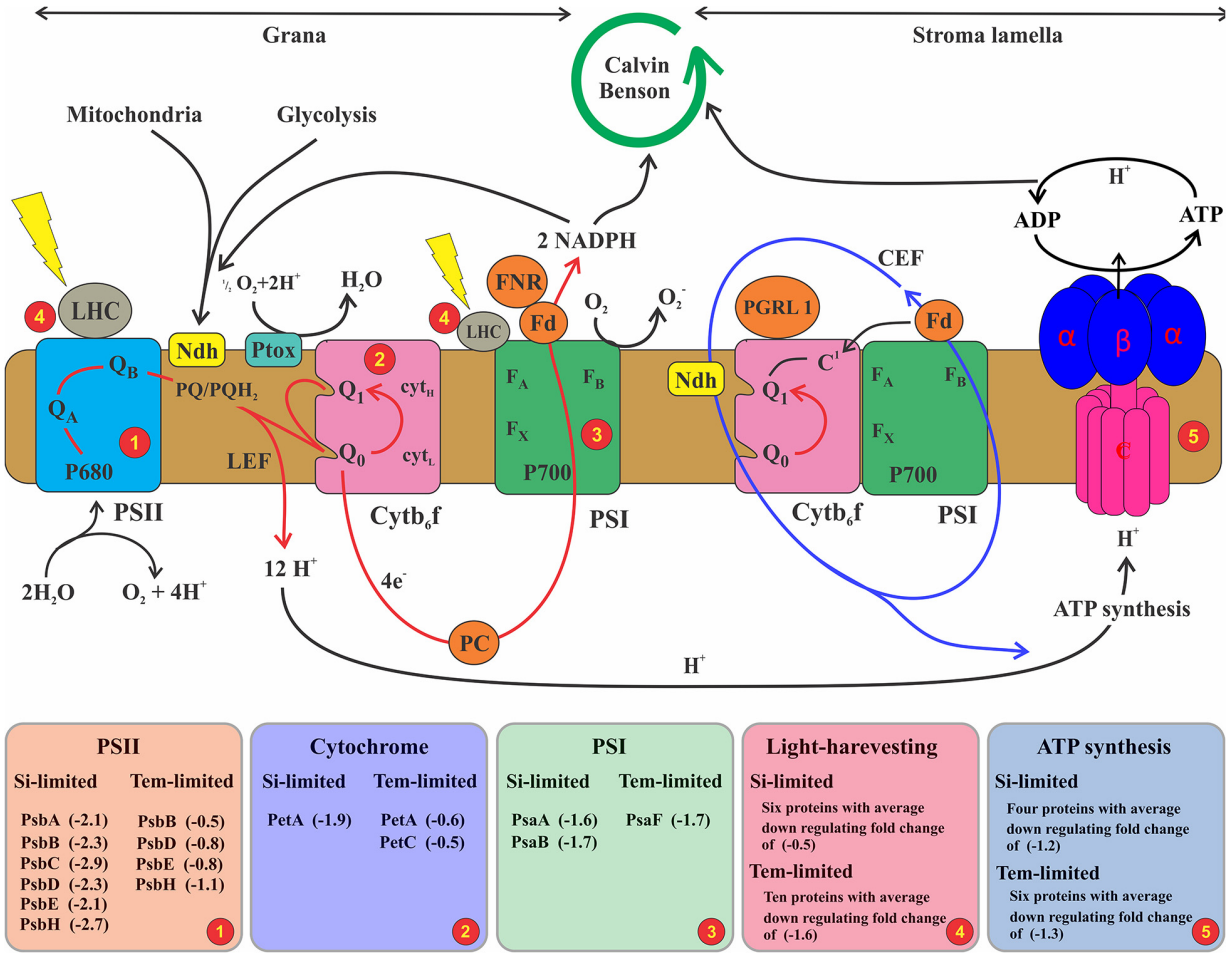


FIGURE 6 | Schematic illustrating the photosynthetic electron transport chain and state transition model. The depicted model consisting of linear and cyclic electron flow and ATP generation. Linear electron flow (red arrow) associated with photosystem II (PSII), cytochrome b6f (Cytb6f), photosystem I (PSI), and ferredoxin-NADPH (FNR). PSII is composed of activation of oxidoreductase transferring an electron from water to plastoquinone, resulting in the reduced form of PQH_2. In this process, when adequate linear electron transfer occurs, it is catalyzed by FNR to NADPH on one side, and simultaneously, proton gradient drives the ATP synthase (ATPase). Cycling electron transport (CEF) (blue lines) is represented by a single pathway involving ferredoxin-PQ reductase, which transfers electrons from Fd to PQ. Further, under aerobic conditions, the plastoquinone terminal (PTOX) could oxidize the PQ pool; therefore, electrons, e.g., from glycolysis, are imported into the chloroplast and lead to a reduction in the PQ pool by an NADH dehydrogenase (Ndh). In this study, the net production of electron conversation by PSII was decreased over a twofold change because of downregulated PSII primary proteins that catalyze this process (see **Supplementary Table 3**). Subsequently, less electron and proton were pumped to the PSI to NADPH and ATP synthase for the Calvin cycle and carbon fixation process. Higher downregulation in LHC 1 by lower temperature (see **Supplementary Table 3**) affects the CEF electron, which has been sent to ATP synthase with less ATP formation for other cellular processes depending on these energies.

defense mechanisms that are controlled at the metabolomic level. For instance, *Phaeodactylum* has a high capacity for the NPQ during temperature fluctuation (Lavaud et al., 2002) through xanthophyll pigment cycling (Goss and Lepetit, 2015) and the light-harvesting proteins (Bailleul et al., 2010). Further, in higher plants, the PSII protein PsbS proved to be an important component for NPQ (Johnson and Ruban, 2010), but no homolog for this gene has been identified in diatoms (Armbrust et al., 2004; Bowler et al., 2008). However, several antenna proteins in diatoms might serve the role of PsbS in photoprotection (Zhu and Green, 2010), suggesting that there might be an interconnection between PSI and LHC to be involved in NPQ of *S. dohrnii* in this study under lower temperature.

AET pathways remove excess reductant from thylakoid membranes (Peltier et al., 2010) to consume excess reducing power generated by photosynthesis, thus decreasing the probability of ROS formation (Niyogi, 2000). The proportion of electrons consumed by AET, measured as light-dependent oxygen consumption, changes with changing growth in *T. pseudonana*, suggesting that AET can also be an important component for maintaining redox balance (Waring et al., 2010) in this study. PSII associated with electron flow, a plastid terminal oxidase, and the transfer of reduced carbon compounds to the mitochondria for oxidation, or the Mehler reaction could be the mechanism of AET in diatoms. AET including the Mehler reaction can consume up to 50% of the electrons released by PSII

in diatoms (Waring et al., 2010). Similarly, an earlier study on *Phaeodactylum* photo physiology under changing environmental conditions predicted that AET could represent an important proportion of the total electron transport in thylakoids (Wagner et al., 2006). Altogether, it appears likely that AET significantly contributed to photoprotection in our experiments, especially Si limitation cells, where PSII proteins decreased dynamically at lower temperatures.

In this study, we did not measure direct AET activities but identified proteins that were involved in this metabolism and were differentially expressed in Si- and temperature-limited cells. For example, NADH dehydrogenase and alternative oxidase (AOX) were downregulated in cells that grew with lower Si, while in lower-temperature cells, these proteins were upregulated. This finding was similar to Thangaraj et al.'s results, in which downregulation of NADH dehydrogenase and AOX response to Si limitation of the diatom *S. dohrnii* were revealed. The NADH dehydrogenase catalyzes oxidation of NADH to NAD^+ and transfers electron to ubiquinone. Upregulation of this NADH dehydrogenase could accelerate the rate of electron transport in the respiratory chain, while downregulation led to decrease its process. The mitochondrial AOX protein used for removing an excess electron in the nutrient-limited diatom (Allen et al., 2008) and upregulation of this protein could be involved in the mitigation of mitochondrial ROS production (Lin et al., 2017) in temperature-limited cells, whereas downregulation of this AOX protein and many other essential enzymes of the FoF1-type proteins in *S. dohrnii* suggests the regulation of ATP production with a coincident blockage of the respiratory chain under Si limitation. Consistent proteomic regulation was seen earlier on diatoms and cyanobacteria responding to a lower temperature (Mackey et al., 2013; Dong et al., 2016) and Si concentration (Chen et al., 2018; Thangaraj et al., 2019). Taken together, in *S. dohrnii* PSII, linear electron flow (grana) to NADPH mechanisms could be regulated by (Si) concentration, whereas PSI, LHC, and cyclic electron flow (stroma) to ATP metabolisms could be modulated by temperature limitation.

Specific Cellular Response in Carbon Metabolism

The carbon and carbohydrate metabolism in diatoms involves complex enzymatic steps or metabolic reactions to convert carbohydrates into metabolic precursors (**Figure 7**) for cellular development; changes in those enzymes could impact the fundamental catalytic information among its intracellular process. In this study, a vital enzyme of GAPD was downregulated in Si-limited cells, while this gene in cells grown at a lower temperature was upregulated. GAPD is a major consumer of NADPH in diatoms; this allows for the redirecting of NADPH to other reductive cellular processes for the ecological success of diatoms under stressful conditions (Mekhalfi et al., 2014). Fumarate and PYK proteins were upregulated in cells grown with lower Si, while these proteins did not change their abundances in lower-temperature cells. Fumarate involves the major steps of the TCA cycle, which can modify protein binding and enzymatic activity (Leshets et al., 2018), while PYK is involved in the lipid biosynthesis of diatoms (Ma et al., 2014). Upregulation of these fumarate and PYK in this study was consistent with their variation in the transcriptome (Bender et al., 2014) and proteome (Smith et al., 2016) response to nitrate and iron deficiency. Further, important amino acids arginine, glutamate, and glutamine were downregulated in Si-limited cells, but they were upregulated in cells with lower temperature. In diatoms, the joint reactions of these enzymes are the primary route of ammonium assimilation and other nitrogenous compounds, providing an important link between ammonium assimilation and carbon metabolism. These results suggested limited nitrogenous assimilation in Si-limited cells compared with enhanced assimilation in lower-temperature cells, supporting the physiological data of inorganic nutrients (**Figure 1**).

Similar to Si limitation, *S. dohrnii* grown at lower temperatures also show distinct proteomic regulation in carbon metabolism. For example, a major intermediate of TCA cycle oxaloacetate (OAA) was downregulated in cells grown at lower temperatures while this protein was upregulated under silicate deficiency. Downregulation of this OAA could inhibit the partial mitochondrial function and the production of precursors for important biomolecules. Further, the key enzyme of phosphofructokinase (pfkA) and PPP proteins (xylulose, sedoheptulose, and erythrose) were upregulated in lower-temperature cells, while these genes did not express in lower-Si cells. The pfka catalyzes the irreversible step in glycolysis, and therefore, upregulation of this enzyme under stress could significantly enhance the glycolytic pathway in diatoms (Yang et al., 2014). Changes in the PPP proteins can operate in different modes: (i) ATP-consuming mode (where there is no loss in carbon or net production of ribulose-5-phosphate as building blocks for other molecules such as AA), (ii) NADPH-producing mode (where all carbon from G-6-P is released as CO2), or (iii) PPP coupled with glycolysis to produce both ATP and NAD(P)H and CO2. These various modes interact and equilibrate with each other when there are energetic needs for the cell, which could favor cellular functions of protein binding and cell cycle regulation under stress conditions. Amino acids of serine and glycine are important for the diatom cell wall formation (Hecky et al., 1973) and were upregulated in lower-temperature cells, while these metabolisms were downregulated in Si-limited cells.

Changes in the Ribosome and Its Associated Metabolism

In general, the aminoacyl-tRNA delivers the amino acid to the ribosome for incorporation into the polypeptide chain and functional proteins that are being produced during the translation process. If the incorrect amino acid is attached during this process, then the tRNA is improperly charged and sends improper amino acids to the ribosome via a translation mechanism that would dynamically impact the ribosome complex and its function. In this study, 49 proteins related to (r, m, and t-RNA) translation mechanisms were downregulated in lower-Si cells, indicating that this mechanism could be nutrient (Si)-dependent in *S. dohrnii*, resulting in 61 downregulated proteins in ribosome metabolism and reduced protein synthesis (**Figure 2**). On the contrary, 29 proteins related to DNA

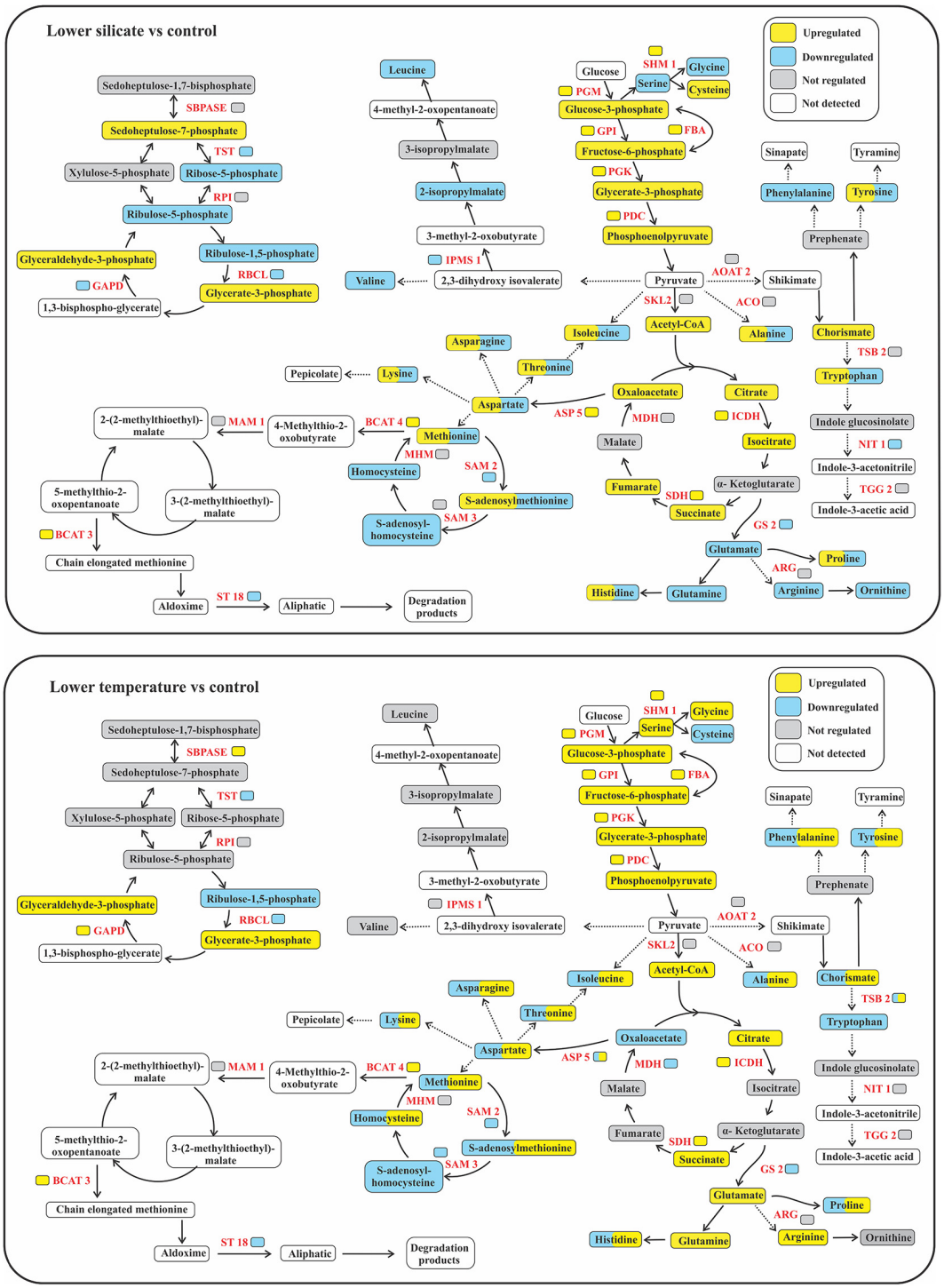


FIGURE 7 | Changes in the protein abundances and carbon flow in the central carbon metabolism of the response of the diatom *S. dohrnii* to silicate and temperature limitation. SBPASE, Sedoheptulose-1,7-bisphosphatase; RPI, ribulose-5-phosphate isomerase; TST, transketolase; RBCL, ribulose bisphosphate carboxylase large chain; GAPD, glyceraldehyde-3-phosphate dehydrogenase; PGM, phosphoglucomutase; SHM, serine hydroxy-methyltransferase; GPI, glucose-6-phosphate isomerase; PGK, phosphoglycerate kinase; FBA, fructose-bisphosphate aldolase; PDC, pyruvate dehydrogenase; AOAT2, alanine-2-oxoglutarate aminotransferase; SKL, shikimate kinase; ACO, aconitate hydratase; ICDH, isocitrate dehydrogenase; SDH, succinate dehydrogenase; MDH, malate dehydrogenase; ASP5, aspartate aminotransferase 5; SAM2, S-adenosylmethionine synthetase 2; SAM3, S-adenosylmethionine synthetase 3; MHM, 5-methyl-tetra hydropteroyltriglutamate-homocysteine methyltransferase; BCAT4, branched-chain amino acid transaminase 4; MAM, methylthioalkylmalate synthase; BCAT3, branched-chain amino acid transaminase 3; IPMS, 2-isopropylmalate synthase; ST18, sulfotransferase 18; TSB2, tryptophan synthase beta-subunit; NIT, nitrilase; TGG2, thioglucoside glucohydrolase 2; GS2, glutamine synthetase.

binding to RNA transcription were downregulated in lower-temperature cells, suggesting that this could be temperature-dependent, resulting in downregulation of only 18 ribosomal proteins. Similarly, recent reports also found that the temperature depended on the transcription processes in *Arabidopsis* (Cortijo et al., 2017) and other eukaryotes (Oliveira et al., 2016).

The primary role of the ribosome is to synthesize functional proteins that are needed for other metabolic functions. These protein synthesis and translation processes are tightly coordinated with cell growth and proliferation (Ruggero and Pandolfi, 2003). Impairment of any of these changes can severely alter cell growth and perturb organism development; this has been demonstrated in almost all eukaryotic organisms (Zhou et al., 2015). For example, a genetic study on yeast revealed that changes in the ribosome assembly result in a lesser amount of protein synthesis and cell size (Jorgensen et al., 2002). Similarly, increasing abundances of elevating ribosome proteins promote cell growth and proliferation by boosting protein synthesis (Van Riggelen et al., 2010). Likewise, several studies have demonstrated in a wide range of organisms that the efficiency of protein synthesis is governed by the rates of ribosome complexes and their components (Choi and Puglisi, 2017; Verma et al., 2019). However, in diatoms, the downregulation of the ribosome not only decreases the material and energy expenditure but also smoothens the polypeptide processing and proteasomes to facilitate the protein turnover under changing Si concentrations (Shrestha and Hildebrand, 2015) and temperatures (Liang et al., 2019). Such protein turnover in diatoms has been reported previously to decrease cell growth, development, proliferation, response to changes in temperature (Liang et al., 2019), CO_2 (Beszteri et al., 2018), and nutrient stress condition (Dyhrman et al., 2012; Chen et al., 2018). These results indicate that if components of a ribosome are increased beyond normal, it would drive efficient protein synthesis, cell growth, and proliferation, while if it decreased, it could inhibit growth and proliferation. It was evident that decreased ribosomal complexes, protein synthesis (**Figure 2**), and cell growth (**Figure 1**) were present in both silicate and temperature limitation.

Common Cellular Responses of *S. dohrnii* to the Lower Silicate and Temperature

In this study, proteins associated with photosynthesis, LHCs, carbon metabolism, and ribosomes were differentially expressed in both Si- and temperature-limited cells. Recent studies show that lower Si and temperature decreased a diatom's protein synthesis and led to metabolic imbalances and oxidative stress with a negative impact on photosynthesis and carbon fixation (Shrestha et al., 2012; Dong et al., 2016; Thangaraj et al., 2019). Among PSII functional proteins, PsbB, PSbD, PsbE, and PsbH were downregulated at both stress conditions, indicating that these proteins are both nutrient (Si)- and temperature-dependent in the marine environment. Similarly, in PSI and cytochrome, PetA and PsbF proteins were also being downregulated at both stress conditions. PetA binds chlorophyll and catalyzes the light-induced photochemical process, while PsbF stimulates electron transfer or the heme-binding or iron-binding process, which could be influenced by both stress conditions. LHC proteins located in the photosystem complex and their increasing enzymatic modulation are dependent on proteins that are rich in N, while their biosynthesis process depends on temperature limitation. Accordingly, lower N acquisition in Si-limited cells and limited catalytic process due to lower temperature (Thangaraj and Sun, 2020a) could share many downregulating LHC proteins in common (**Figure 6**). These results are consistent with the response of model diatoms *T. pseudonana* (Chen et al., 2018) and *P. tricornutum* (Yang et al., 2014) to nutrient limitation, suggesting that these both stress Si and temperature conditions could worsen light reactions of *S. dohrnii*.

Carbon fixation proteins: Phosphoribulokinase (PPDK), CA, and RuBisCO were downregulated in both stressed cells (**Figure 6**), suggesting that both Si and temperature limitation could reduce the C4-like carbon concentrating pathway and decrease carbon fixation, influencing the carbon inside the cell to be adjusted to the match reductant supply (Allen et al., 2008). Similarly, S-adenosylmethionine synthetase (SAM) was downregulated at both conditions, which is a storage form of methionine (Bourgis et al., 1999) that is involved in over 40 metabolic functions/reactions associated with nucleic acids, lipids, proteins, and secondary metabolites (Bartlem et al., 2000). Therefore, changes in this SAM in cells grown at stressed conditions could inhibit cellular development and growth (Reintanz et al., 2001). Besides, TST is a key enzyme involved in sugar metabolism and was downregulated at both stress conditions. In algae, TST is located in the plastid membrane, where photosynthesis occurs (Teige et al., 1998); thus, changes in TST could impact the photosynthesis process (Henkes et al., 2001). Despite different fold changes, isocitrate (ICDH) and succinate (SDH) were also being downregulated in both conditions. ICDH is a vital enzyme in the two-step process of producing alpha-ketoglutarate and CO_2 in diatoms (Alipanah et al., 2015), while SDH involved in the mitochondrial electron transport chain for transferring electrons from succinate to ubiquinone could be influenced by both stress conditions.

Chen et al. (2018) recently reported that, although diatom carbon fixation was decreased during stress conditions, cellular C and lipid content was increased due to the upregulation of the glycolytic metabolic pathway. Their findings are also supported in the present study: upregulation of glycolytic pathways (**Figure 6**) resulted in enhancement in cellular C and lipid accumulation under both stress conditions (**Figure 2**). In addition, under both conditions, acetyl-CoA carboxylase (ACACA) was upregulated, which is responsible for the production of long-chain acyl-CoA that can be utilized for the cellular lipid synthesis and variant subcellular localization to different anabolic and catabolic pathways (Mashek et al., 2007). Further, identified upregulated glycolytic proteins in both stress conditions may be the response to decreased production of NADPH and ATP from photosynthetic light reactions. Alternatively, increased sugar (glucose, sucrose, and fructose) content could suppress the expression of light-harvesting proteins (Martin et al., 2011) and light reactions (Chen et al., 2005). This is consistent with the observation of the negative relationship between the photosynthetic rate and carbohydrate content (Araya et al., 2010). It is likely that the downregulation of aliphatic

glucosinolates causes increased carbohydrate accumulation and imbalances between light and dark reactions, which, in turn, eventually decreases photosynthesis and slow growth rate at both stress conditions.

In addition, both stress conditions caused upregulation of the methionine chain-elongation (BCAT) protein, catalyzing the terminal steps of methionine, which could be a substrate enhancement for cellular development. In connection to this, S-adenosylmethionine was downregulated, suggesting that channeling methionine was inhibited to form SAM for one-carbon metabolism when this methionine pathway was regulated by both conditions. The result agrees with the mutation of SAM that can result in a dramatic accumulation of methionine (Shen et al., 2002) for cellular development. In addition, many amino acids in both conditions were expressed similarly to up- and downregulation, respectively. These included multiple aminotransferases, which yield fates for amino acids including the rearrangement of new amino acid formation or complete intracellular recycling to alpha, keto acids, ammonia, or pyruvate. The results show that, despite photosynthesis, electron transport, light-harvesting, and the carbon fixation processes being decreased by both conditions, the total lipid and cellular carbon content per cell increased due to the utilization of upregulated proteins in glycolytic pathways and long-chain fatty acids.

Histones are a primary component of eukaryotic chromatin; modulation of this can affect the DNA binding, replication, and protein folding relevant to cell cycle progress (Li et al., 2018). It is shown that decreased abundances of histone in both conditions in this study would greatly impact smooth cell cycle progress and cellular development of diatoms. The MMR protein and associated DNA repair are essential to maintain the integrity and continuity of general information in eukaryotic cells. The study shows downregulation of this MMR at both stress conditions, suggesting a reduction in the DNA repair process. PCNA is a potential molecular marker for the phytoplankton metabolic process, and its higher expression ensures a smooth cell cycle (Zhang et al., 2019). Consistent with this, decreased abundances of this PCNA in this study indicate a further reduction in the cell cycle process of *S. dohrnii.* Cell division CDK is an engine of the cell cycle in phytoplankton and their DNA replication (Zhang et al., 2019). Downregulation of this CDK at both conditions shows that it might play an essential role in decreasing cell growth and abundances under stress conditions like Si and temperature.

CONCLUSION

The results showed that the limitation of both silicate and temperature has common and specific metabolic responses to inhibit cellular development and growth of the diatom *S. dohrnii.* Nutrient assimilation and utilization were decreased at both conditions at different extent, reflected to impact PSII and NADPH production in silicate-limited cells and PSI and ATP production in lower-temperature cells. Notably, both stress depressed carbon fixation and photorespiration but simultaneously endorsed lipid and carbohydrate accumulation. Besides, both limitations have their own cell cycle-related protein alteration leading to metabolic imbalances in the transcription and translation process in temperature- and silicate-limited cells, respectively. Moreover, ribosome assembly in both limitations decreased its process, although in different extent, leading to reduced functional protein synthesis for the cellular development and growth of the diatom *S. dohrnii.* The integrative approach revealed previously unrecognized silicate and temperature that are dependent on mechanisms in growth and cellular development of marine diatoms, which could be valuable to understand how nutrients (Si) and the temperature are different for a diatom's functional activity in the marine environment for their ecological success.

DATA AVAILABILITY STATEMENT

The raw mass spectrometry proteomics data and analysis file have been submitted to ProteomeXchange via PRIDE database (www.ebi.ac.uk/pride/archive/) with identifier PXD021705.

AUTHOR CONTRIBUTIONS

ST and JS designed the study. ST performed the laboratory experiment, carried out data analysis, and defined the manuscript content. MG reviewed and revised the manuscript. JS coordinated this investigation and provided guidance, funding, and facilities to perform the experiment. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by the National Key Research and Development Project of China (2019YFC1407805), the National Natural Science Foundation of China (41876134, 41676112, and 41276124), the Tianjin 131 Innovation Team Program (20180314), and the Changjiang Scholar Program of Chinese Ministry of Education (T2014253) to JS.

ACKNOWLEDGMENTS

JS dedicate this article exclusively to the tribute of deceased MG for his significant contribution to the field of Algal Biology.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.578915/full#supplementary-material

Supplementary Table 1 | Annotation details of overall identified proteins.

Supplementary Table 2 | Identification of down and up-regulated proteins in silicate and temperature limited conditions.

Supplementary Table 3 | DEP associated with photosynthesis, carbon and ribosome metabolism in both silicate and temperature limited conditions.

REFERENCES

Alipanah, L., Rohloff, J., Winge, P., Bones, A. M., and Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 66, 6281–6296.

Allen, A. E., Dupont, C. L., Obornik, M., Horak, A., Nunes-Nesi, A., Mccrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008). Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10438–10443. doi: 10.1073/pnas.0711370105

Araya, T., Noguchi, K., and Terashima, I. (2010). Effect of nitrogen nutrition on the carbohydrate repression of photosynthesis in leaves of *Phaseolus vulgaris* L. *J. Plant Res.* 123, 371–379. doi: 10.1007/s10265-009-0279-8

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306, 79–86. doi: 10.1126/science.1101156

Bailleul, B., Rogato, A., De Martino, A., Coesel, S., Cardol, P., Bowler, C., et al. (2010). An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18214–18219. doi: 10.1073/pnas.1007703107

Bartlem, D., Lambein, I., Okamoto, T., Itaya, A., Uda, Y., Kijima, F., et al. (2000). Mutation in the threonine synthase gene results in an over-accumulation of soluble methionine in Arabidopsis. *Plant Physiol.* 123, 101–110. doi: 10.1104/pp.123.1.101

Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L., and Armbrust, E. (2014). Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* 1:3. doi: 10.3389/fmars.2014.00003

Bermudez, R., Feng, Y., Roleda, M. Y., Tatters, A. O., Hutchins, D. A., Larsen, T., et al. (2015). Long-term conditioning to elevated pCO2 and warming influences the fatty and amino acid composition of the diatom *Cylindrotheca* fusiformis. *PLoS One* 10:e0123945. doi: 10.1371/journal.pone.0123945

Beszteri, S., Thoms, S., Benes, V., Harms, L., and Trimborn, S. (2018). The response of three southern ocean phytoplankton species to ocean acidification and light availability: a transcriptomic study. *Protist* 169, 958–975. doi: 10.1016/j.protis.2018.08.003

Bourgis, F., Roje, S., Nuccio, M. L., Fisher, D. B., Tarczynski, M. C., Li, C., et al. (1999). S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* 11, 1485–1497. doi: 10.2307/3870977

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244.

Boyd, P. W., Dillingham, P. W., Mcgraw, C. M., Armstrong, E. A., Cornwall, C. E., Feng, Y. Y., et al. (2015). Physiological responses of a Southern Ocean diatom to complex future ocean conditions. *Nat. Clim. Chang.* 6, 207–213. doi: 10.1038/nclimate2811

Boyd, P. W., and Hutchins, D. A. (2012). Understanding the responses of ocean biota to a complex matrix of cumulative anthropogenic change. *Mar. Ecol. Prog. Ser.* 470, 125–135. doi: 10.3354/meps10121

Brembu, T., Muhlroth, A., Alipanah, L., and Bones, A. M. (2017). The effects of phosphorus limitation on carbon metabolism in diatoms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160406. doi: 10.1098/rstb.2016.0406

Brunet, C., and Lavaud, J. (2010). Can the xanthophyll cycle help extract the essence of the microalgal functional response to a variable light environment? *J. Plankton Res.* 32, 1609–1617. doi: 10.1093/plankt/fbq104

Bucciarelli, E., and Sunda, W. G. (2003). Influence of CO2, nitrate, phosphate, and silicate limitation on intracellular dimethylsulfoniopropionate in batch cultures of the coastal diatom *Thalassiosira pseudonana*. *Limnol. Oceanogr.* 48, 2256–2265. doi: 10.4319/lo.2003.48.6.2256

Charbonneau, M. E., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., et al. (2007). O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *J Bacteriol.* 189, 8880–8889. doi: 10.1128/jb.00969-07

Chen, G.-Y., Yong, Z.-H., Liao, Y., Zhang, D.-Y., Chen, Y., Zhang, H.-B., et al. (2005). Photosynthetic acclimation in rice leaves to free-air CO2 enrichment related to both ribulose-1, 5-bisphosphate carboxylation limitation and ribulose-1, 5-bisphosphate regeneration limitation. *Plant Cell Physiol.* 46, 1036–1045.

Chen, X. H., Li, Y. Y., Zhang, H., Liu, J. L., Xie, Z. X., Lin, L., et al. (2018). Quantitative proteomics reveals common and specific responses of a marine diatom *Thalassiosira pseudonana* to different macronutrient deficiencies. *Front. Microbiol.* 9:2761. doi: 10.3389/fmicb.2018.02761

Choi, J., and Puglisi, J. D. (2017). Three tRNAs on the ribosome slow translation elongation. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13691–13696. doi: 10.1073/pnas.1719592115

Cortijo, S., Charoensawan, V., Brestovitsky, A., Buning, R., Ravarani, C., Rhodes, D., et al. (2017). Transcriptional regulation of the ambient temperature response by H2A. Z nucleosomes and HSF1 transcription factors in Arabidopsis. *Mol. Plant* 10, 1258–1273. doi: 10.1016/j.molp.2017.08.014

Depauw, F. A., Rogato, A., Ribera D'alcalá, M., and Falciatore, A. (2012). Exploring the molecular basis of responses to light in marine diatoms. *J. Exp. Bot.* 63, 1575–1591. doi: 10.1093/jxb/ers005

Dong, H. P., Dong, Y. L., Cui, L., Balamurugan, S., Gao, J., Lu, S. H., et al. (2016). High light stress triggers distinct proteomic responses in the marine diatom *Thalassiosira pseudonana*. *BMC Genomics* 17:994. doi: 10.1186/s12864-016-3335-5

Du, C., Liang, J.-R., Chen, D.-D., Xu, B., Zhuo, W.-H., Gao, Y.-H., et al. (2014). iTRAQ-based proteomic analysis of the metabolism mechanism associated with silicon response in the marine diatom *Thalassiosira pseudonana*. *J. Proteome Res.* 13, 720–734. doi: 10.1021/pr400803w

Durkin, C. A., Koester, J. A., Bender, S. J., and Armbrust, E. V. (2016). The evolution of silicon transporters in diatoms. *J. Phycol.* 52, 716–731. doi: 10.1111/jpy.12441

Dyhrman, S. T., Jenkins, B. D., Rynearson, T. A., Saito, M. A., Mercier, M. L., Alexander, H., et al. (2012). The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS One* 7:e33768. doi: 10.1371/journal.pone.0033768

Edmundson, S. J., and Huesemann, M. H. (2015). The dark side of algae cultivation: characterizing night biomass loss in three photosynthetic algae, *Chlorella sorokiniana*, *Nannochloropsis salina* and *Picochlorum* sp. *Algal Res.* 12, 470–476. doi: 10.1016/j.algal.2015.10.012

Falkowski, P. G., Barber, R. T., and Smetacek, V. (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* 281, 200–206. doi: 10.1126/science.281.5374.200

Goss, R., and Lepetit, B. (2015). Biodiversity of NPQ. *J. Plant Physiol.* 172, 13–32. doi: 10.1016/j.jplph.2014.03.004

Guerra, L. T., Levitan, O., Frada, M. J., Sun, J. S., Falkowski, P. G., and Dismukes, G. C. (2013). Regulatory branch points affecting protein and lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. *Biomass Bioenergy* 59, 306–315. doi: 10.1016/j.biombioe.2013.10.007

Hamm, C. E., Merkel, R., Springer, O., Jurkojc, P., Maier, C., Prechtel, K., et al. (2003). Architecture and material properties of diatom shells provide effective mechanical protection. *Nature* 421, 841–843. doi: 10.1038/nature01416

Hecky, R., Mopper, K., Kilham, P., and Degens, E. (1973). The amino acid and sugar composition of diatom cell-walls. *Mar. Biol.* 19, 323–331. doi: 10.1007/bf00348902

Henkes, S., Sonnewald, U., Badur, R., Flachmann, R., and Stitt, M. (2001). A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *Plant Cell* 13, 535–551. doi: 10.2307/3871405

Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 158, 299–312. doi: 10.1104/pp.111.184333

Howden, A. J., and Preston, G. M. (2009). Nitrilase enzymes and their role in plant-microbe interactions. *Microb. Biotechnol.* 2, 441–451. doi: 10.1111/j.1751-7915.2009.00111.x

Huysman, M. J., Fortunato, A. E., Matthijs, M., Costa, B. S., Vanderhaeghen, R., Van Den Daele, H., et al. (2013). AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). *Plant Cell* 25, 215–228. doi: 10.1105/tpc.112.106377

Jallet, D., Caballero, M. A., Gallina, A. A., Youngblood, M., and Peers, G. (2016). Photosynthetic physiology and biomass partitioning in the model diatom

Phaeodactylum tricornutum grown in a sinusoidal light regime. *Algal Res.* 18, 51–60. doi: 10.1016/j.algal.2016.05.014

Jeffrey, S. W., and Humphrey, G. F. (1975). New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. *Biochem. Physiol. Pflanz.* 167, 191–194. doi: 10.1016/s0015-3796(17)30778-3

Jian, J., Zeng, D., Wei, W., Lin, H., Li, P., and Liu, W. (2017). The combination of RNA and protein profiling reveals the response to nitrogen depletion in *Thalassiosira pseudonana*. *Sci. Rep.* 7:8989.

Johnson, M. P., and Ruban, A. V. (2010). Arabidopsis plants lacking PsbS protein possess photoprotective energy dissipation. *Plant J.* 61, 283–289. doi: 10.1111/j.1365-313x.2009.04051.x

Jorgensen, P., Nishikawa, J. L., Breitkreutz, B.-J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. *Science* 297, 395–400. doi: 10.1126/science.1070850

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070

Kruger, N. J. (2009). "The bradford method for protein quantitation," in *The Protein Protocols Handbook*, ed. J. M. Walker (Totowa, NJ: Springer), 17–24.

Lavaud, J., Rousseau, B., Van Gorkom, H. J., and Etienne, A.-L. (2002). Influence of the diadinoxanthin pool size on photoprotection in the marine planktonic diatom *Phaeodactylum tricornutum*. *Plant Physiol.* 129, 1398–1406. doi: 10.1104/pp.002014

Leshets, M., Silas, Y. B., Lehming, N., and Pines, O. (2018). Fumarase: from the TCA Cycle to DNA damage response and tumor suppression. *Front. Mol. Biosci.* 5:68. doi: 10.3389/fmolb.2018.00068

Li, X., Egervari, G., Wang, Y., Berger, S. L., and Lu, Z. (2018). Regulation of chromatin and gene expression by metabolic enzymes and metabolites. *Nat. Rev. Mol. Cell Biol.* 19, 563–578. doi: 10.1038/s41580-018-0029-7

Liang, Y., Koester, J. A., Liefer, J. D., Irwin, A. J., and Finkel, Z. V. (2019). Molecular mechanisms of temperature acclimation and adaptation in marine diatoms. *ISME J.* 13, 2415–2425. doi: 10.1038/s41396-019-0441-9

Lin, Q., Liang, J.-R., Huang, Q.-Q., Luo, C.-S., Anderson, D. M., Bowler, C., et al. (2017). Differential cellular responses associated with oxidative stress and cell fate decision under nitrate and phosphate limitations in *Thalassiosira pseudonana*: comparative proteomics. *PLoS One* 12:e0184849. doi: 10.1371/journal.pone.0184849

Lott, B. B., Wang, Y., and Nakazato, T. (2013). A comparative study of ribosomal proteins: linkage between amino acid distribution and ribosomal assembly. *BMC Biophys.* 6:13. doi: 10.1186/2046-1682-6-13

Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* 193, 265–275.

Ma, Y.-H., Wang, X., Niu, Y.-F., Yang, Z.-K., Zhang, M.-H., Wang, Z.-M., et al. (2014). Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom *Phaeodactylum tricornutum*. *Microb. Cell Fact.* 13, 1–9. doi: 10.1007/978-94-007-7864-1_127-1

Mackey, K. R., Paytan, A., Caldeira, K., Grossman, A. R., Moran, D., Mcilvin, M., et al. (2013). Effect of temperature on photosynthesis and growth in marine *Synechococcus* spp. *Plant Physiol.* 163, 815–829. doi: 10.1104/pp.113.221937

Maranon, E., Lorenzo, M. P., Cermeno, P., and Mourino-Carballido, B. (2018). Nutrient limitation suppresses the temperature dependence of phytoplankton metabolic rates. *ISME J.* 12, 1836–1845. doi: 10.1038/s41396-018-0105-1

Martin, P., Van Mooy, B. A., Heithoff, A., and Dyhrman, S. T. (2011). Phosphorus supply drives rapid turnover of membrane phospholipids in the diatom *Thalassiosira pseudonana*. *ISME J.* 5, 1057–1060. doi: 10.1038/ismej.2010.192

Mashek, D. G., Li, L. O., and Coleman, R. A. (2007). Long-chain acyl-CoA synthetases and fatty acid channeling. *Future Lipidol.* 2, 465–476. doi: 10.2217/17460875.2.4.465

Mekhalfi, M., Puppo, C., Avilan, L., Lebrun, R., Mansuelle, P., Maberly, S. C., et al. (2014). Glyceraldehyde-3-phosphate dehydrogenase is regulated by ferredoxin-NADP reductase in the diatom A sterionella formosa. *New Phytol.* 203, 414–423. doi: 10.1111/nph.12820

Miflin, B. J., and Habash, D. Z. (2002). The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. *J. Exp. Bot.* 53, 979–987. doi: 10.1093/jexbot/53.370.979

Niyogi, K. K. (2000). Safety valves for photosynthesis. *Curr. Opin. Plant Biol.* 3, 455–460. doi: 10.1016/s1369-5266(00)00113-8

Nunn, B. L., Faux, J. F., Hippmann, A. A., Maldonado, M. T., Harvey, H. R., Goodlett, D. R., et al. (2013). Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLoS One* 8:e75653. doi: 10.1371/journal.pone.0075653

Oliveira, S. M., Häkkinen, A., Lloyd-Price, J., Tran, H., Kandavalli, V., and Ribeiro, A. S. (2016). Temperature-dependent model of multi-step transcription initiation in *Escherichia coli* based on live single-cell measurements. *PLoS Comput. Biol.* 12:e1005174. doi: 10.1371/journal.pcbi.1005174

Peltier, G., Tolleter, D., Billon, E., and Cournac, L. (2010). Auxiliary electron transport pathways in chloroplasts of microalgae. *Photosynth. Res.* 106, 19–31. doi: 10.1007/s11120-010-9575-3

Reintanz, B., Lehnen, M., Reichelt, M., Gershenzon, J., Kowalczyk, M., Sandberg, G., et al. (2001). bus, a bushy Arabidopsis CYP79F1 knockout mutant with abolished synthesis of short-chain aliphatic glucosinolates. *Plant Cell* 13, 351–367. doi: 10.2307/3871281

Ruggero, D., and Pandolfi, P. P. (2003). Does the ribosome translate cancer? *Nat. Rev. Cancer* 3, 179–192. doi: 10.1038/nrc1015

Secco, D., Wang, C., Arpat, B. A., Wang, Z., Poirier, Y., Tyerman, S. D., et al. (2012). The emerging importance of the SPX domain-containing proteins in phosphate homeostasis. *New Phytol.* 193, 842–851. doi: 10.1111/j.1469-8137.2011.04002.x

Shen, B., Li, C., and Tarczynski, M. C. (2002). High free-methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl-L-methionine synthetase 3 gene. *Plant J.* 29, 371–380. doi: 10.1046/j.1365-313x.2002.01221.x

Shrestha, R. P., and Hildebrand, M. (2015). Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryot. Cell* 14, 29–40. doi: 10.1128/ec.00209-14

Shrestha, R. P., Tesson, B., Norden-Krichmar, T., Federowicz, S., Hildebrand, M., and Allen, A. E. (2012). Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* 13:499. doi: 10.1186/1471-2164-13-499

Smith, S. R., Gle, C., Abbriano, R. M., Traller, J. C., Davis, A., Trentacoste, E., et al. (2016). Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytol.* 210, 890–904. doi: 10.1111/nph.13843

Sunda, W. G., Price, N. M., and Morel, F. M. (2005). Trace metal ion buffers and their use in culture studies. *Algal Culturing Tech.* 4, 35–63. doi: 10.1016/b978-012088426-1/50005-6

Takahashi, H., Kopriva, S., Giordano, M., Saito, K., and Hell, R. (2011). Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes. *Annu. Rev. Plant Biol.* 62, 157–184. doi: 10.1146/annurev-arplant-042110-103921

Teige, M., Melzer, M., and Süss, K. H. (1998). Purification, properties and in situ localization of the amphibolic enzymes D-ribulose 5-phosphate 3-epimerase and transketolase from spinach chloroplasts. *Eur. J. Biochem.* 252, 237–244. doi: 10.1046/j.1432-1327.1998.2520237.x

Thangaraj, S., Shang, X., Sun, J., and Liu, H. (2019). Quantitative proteomic analysis reveals novel insights into intracellular silicate stress-responsive mechanisms in the diatom *Skeletonema dohrnii*. *Int. J. Mol. Sci.* 20:2540. doi: 10.3390/ijms20102540

Thangaraj, S., and Sun, J. (2020a). Transcriptomic reprogramming of the oceanic diatom *Skeletonema dohrnii* under warming ocean and acidification. *Environ. Microbiol.* doi: 10.1111/1462-2920.15248

Thangaraj, S., and Sun, J. (2020b). The biotechnological potential of the marine diatom *Skeletonema dohrnii* to the elevated temperature and pCO2 concentration. *Mar. Drugs* 18, 259. doi: 10.3390/md18050259

Unwin, R. D. (2010). Quantification of proteins by iTRAQ. *Methods Mol. Biol.* 658, 205–215. doi: 10.1007/978-1-60761-780-8_12

Van Mooy, B. A., Fredricks, H. F., Pedler, B. E., Dyhrman, S. T., Karl, D. M., Koblizek, M., et al. (2009). Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* 458, 69–72. doi: 10.1038/nature07659

Van Riggelen, J., Yetil, A., and Felsher, D. W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat. Rev. Cancer* 10, 301–309. doi: 10.1038/nrc2819

Verma, M., Choi, J., Cottrell, K. A., Lavagnino, Z., Thomas, E. N., Pavlovic-Djuranovic, S., et al. (2019). A short translational ramp determines the efficiency of protein synthesis. *Nat. Commun.* 10:5774.

Wagner, H., Jakob, T., and Wilhelm, C. (2006). Balancing the energy flow from captured light to biomass under fluctuating light conditions. *New Phytol.* 169, 95–108. doi: 10.1111/j.1469-8137.2005.01550.x

Waring, J., Klenell, M., Bechtold, U., Underwood, G. J., and Baker, N. R. (2010). Light-induced responses of oxygen phororeduction, reactive oxygen species production and scavenging in two diatom species 1. *J. Phycol.* 46, 1206–1217. doi: 10.1111/j.1529-8817.2010.00919.x

Yang, Z. K., Zheng, J. W., Niu, Y. F., Yang, W. D., Liu, J. S., and Li, H. Y. (2014). Systems-level analysis of the metabolic responses of the diatom *Phaeodactylum tricornutum* to phosphorus stress. *Environ. Microbiol.* 16, 1793–1807. doi: 10.1111/1462-2920.12411

Yoneda, K., Yoshida, M., Suzuki, I., and Watanabe, M. M. (2018). Homologous expression of lipid droplet protein-enhanced neutral lipid accumulation in the marine diatom *Phaeodactylum tricornutum*. *J. Appl. Phycol.* 30, 2793–2802. doi: 10.1007/s10811-018-1402-9

Zhang, H., Liu, J., He, Y., Xie, Z., Zhang, S., Zhang, Y., et al. (2019). Quantitative proteomics reveals the key molecular events occurring at different cell cycle phases of the in situ blooming dinoflagellate cells. *Sci Total Environ.* 676, 62–71. doi: 10.1016/j.scitotenv.2019.04.216

Zhang, S.-F., Yuan, C.-J., Chen, Y., Chen, X.-H., Li, D.-X., Liu, J.-L., et al. (2016). Comparative transcriptomic analysis reveals novel insights into the adaptive response of *Skeletonema costatum* to changing ambient phosphorus. *Front. Microbiol.* 7:1476. doi: 10.3389/fmicb.2016.01476

Zhou, X., Liao, W.-J., Liao, J.-M., Liao, P., and Lu, H. (2015). Ribosomal proteins: functions beyond the ribosome. *J. Mol. Cell Biol.* 7, 92–104. doi: 10.1093/jmcb/mjv014

Zhu, S.-H., and Green, B. R. (2010). Photoprotection in the diatom *Thalassiosira pseudonana*: role of LI818-like proteins in response to high light stress. *Biochim. Biophys. Acta* 1797, 1449–1457. doi: 10.1016/j.bbabio.2010.04.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Thangaraj, Giordano and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

MINI REVIEW
published: 17 December 2020
doi: 10.3389/fpls.2020.589026

Amino Acid Catabolism During Nitrogen Limitation in *Phaeodactylum tricornutum*

*Yufang Pan[1], Fan Hu[2], Chen Yu[1,3], Chenjie Li[1,3], Teng Huang[1,3] and Hanhua Hu[1]**

[1] Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China, [2] School of Foreign Languages, China University of Geosciences, Wuhan, China, [3] University of Chinese Academy of Sciences, Beijing, China

Diatoms can accumulate high levels of triacylglycerols (TAGs) under nitrogen depletion and have attracted increasing attention as a potential system for biofuel production. In *Phaeodactylum tricornutum*, a model diatom, about 40% of lipid is synthesized from the breakdown of cellular components under nitrogen starvation. Our previous studies indicated that carbon skeletons from enhanced branched-chain amino acid (BCAA) degradation under nitrogen deficiency contribute to TAG biosynthesis in *P. tricornutum*. In this review, we outlined the catabolic pathways of all 20 amino acids based on the genome, transcriptome, proteome, and metabolome data. The contribution of these amino acid catabolic pathways to TAG accumulation was also analyzed.

Keywords: diatom, amino acid, catabolism, nitrogen, triacylglycerols

OPEN ACCESS

Edited by:
Miroslav Obornik, Institute of Parasitology, Academy of Sciences of the Czech Republic (ASCR), Czechia

Reviewed by:
Jeroen T. F. Gillard, California State University, Bakersfield, United States
Justine Marchand, Le Mans Université, France

****Correspondence:***
Hanhua Hu
hanhuahu@ihb.ac.cn

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *07 August 2020*
Accepted: *23 November 2020*
Published: *17 December 2020*

Citation:
Pan Y, Hu F, Yu C, Li C, Huang T and Hu H (2020) Amino Acid Catabolism During Nitrogen Limitation in Phaeodactylum tricornutum. Front. Plant Sci. 11:589026. doi: 10.3389/fpls.2020.589026

INTRODUCTION

Diatoms are a group of unicellular eukaryotic algae and an important component of marine phytoplankton. Although thousands of genes were of green algal derivation (Moustafa et al., 2009), diatoms are believed to emerge as the result of a secondary endosymbiotic event between two eukaryotes, a red alga and an oomycete (Medlin et al., 2000). Therefore, diatoms possess some unique features in comparison with other photosynthetic eukaryotes, including the presence of hundreds of genes from bacteria, the Entner–Doudoroff pathway, and urea cycle unfound in plants and green algae (Bowler et al., 2008; Allen et al., 2011; Fabris et al., 2012; Singh et al., 2015). *Phaeodactylum tricornutum* is one of the model diatoms with short generation time, and routine and simple genetic manipulation is available (Zhang and Hu, 2014; Karas et al., 2015; Falciatore et al., 2020). This diatom has the capacity to accumulate eicosapentaenoic acid, fucoxanthin, and neutral lipids (mostly triacylglycerols, TAGs) and thus is perceived as a microalgal cell factory and a potential system for biofuel production (Butler et al., 2020).

Nitrogen, accounting for over 7% of cellular mass in marine microorganisms, is one of the major constituents of both proteins and nucleic acids (Geider and La Roche, 2002). Many nutrients and nitrogen in particular are restricted in the open ocean (Moore et al., 2013), though seasonal inputs of nitrate (NO_3^-) can cause diatom-dominated phytoplankton blooms in coastal ecosystems (Kudela and Dugdale, 2000). Diatoms are capable of assimilating dissolved nitrogen sources of different forms, including inorganic ones such as nitrate (NO_3^-), nitrite (NO_2^-), and ammonium (NH_4^+) and organic ones such as urea and amino acids (Jauffrais et al., 2016). Amino acids can be taken up by cells and intracellularly metabolized as diatom genomes containing plasma membrane amino acid transporters (Armbrust et al., 2004; Sipler and Bronk, 2015), and they can also be oxidized by extracellular L-amino acid oxidase to produce α-keto acid, NH_4^+, and

hydrogen peroxide (Palenik and Morel, 1990; Rees and Allison, 2006; Contreras and Gillard, 2020). Under nitrogen stress, cellular protein content decreases and amino acid degradation occurs (Guerra et al., 2013). Acetyl-CoA, a product of the metabolism of some amino acids, enters the tricarboxylic acid (TCA) cycle and is shunted toward fatty acid biosynthesis (Hockin et al., 2012; Ge et al., 2014; Levitan et al., 2015). It is indicated that carbon skeletons from enhanced branched-chain amino acid (BCAA) degradation under nitrogen deficiency feed into the TCA cycle and contribute to TAG biosynthesis in *P. tricornutum* (Ge et al., 2014; Pan et al., 2017) and *Chlamydomonas reinhardtii* (Liang et al., 2019). However, few studies have examined the interaction between metabolic pathways of the other amino acids and TAG biosynthesis. Although amino acid biosynthesis pathways have been reviewed (Bromke, 2013), catabolic pathways of amino acids in *P. tricornutum* have not been summarized. In plants, amino acid catabolism and regulation have received considerable attention (Hildebrandt et al., 2015), and amino acid catabolism is important not only during normal senescence but also in stress tolerance. The capacity of diatoms to use dissolved amino acids has been considered to help diatoms survive in blooms or in light-impenetrable sediments (Admiraal and Peletier, 1979). In response to nitrogen deprivation, amino acid degradation could promote the redistribution of carbon and nitrogen flow in diatom cells (Alipanah et al., 2015).

In this review, we outlined the catabolic pathways of all 20 amino acids and provided the subcellular localization information of related enzymes according to the prediction (**Supplementary Material**) from genome annotation in *P. tricornutum*. Based on published transcriptomes (Levitan et al., 2015; Matthijs et al., 2016, 2017; Remmers et al., 2018; Smith et al., 2019), proteomes (Remmers et al., 2018), and metabolomes (Ge et al., 2014), we arranged the expression levels of related enzymes and the content of amino acids to interpret the contribution of amino acid degradation to TAG accumulation. In addition, the homologous genes and their transcription levels (Bender et al., 2014) involved in the catabolic pathways of amino acids in *Thalassiosira pseudonana* were also provided.

LEUCINE, ISOLEUCINE, AND VALINE

Branched-chain amino acid (leucine, valine, and isoleucine) content decreased in *P. tricornutum* cells during TAG accumulation (**Supplementary Figure 1**). The catabolism of BCAAs has been mostly unraveled in our previous studies (**Figures 1A,B**; Ge et al., 2014; Pan et al., 2017). The initial steps in the degradation pathways of BCAAs are catalyzed by branched-chain amino acid transaminase (BCAT), which also catalyzes the final step in BCAA synthesis (**Figures 1A,B**). Many copies of BCATs have been identified in both plants and humans. In *Arabidopsis*, there are seven isoforms of BCAT localized in different compartments, and the mitochondrial isoform BCAT2 has been shown to be especially relevant to degradation (Angelovici et al., 2013). Six *BCATs* were annotated in *P. tricornutum* genome, and all *BCATs* could be up-regulated during nitrogen limitation except *BCAT1* (**Figure 2A** and **Supplementary Table 1**). The function of BCATs depends on their localizations (Campbell et al., 2001), and thus, it is reasonable that the chloroplast-localized *P. tricornutum* BCAT1, which may be mainly responsible for BCAA synthesis, is down-regulated during nitrogen limitation. Transcriptional and/or protein levels of the other genes involved in BCAA catabolism were also found to be up-regulated during nitrogen limitation in our previous studies (Ge et al., 2014; Pan et al., 2017). The roles of methylcrotonyl-CoA carboxylase (MCC), propionyl-CoA carboxylase (PCC), 3-hydroxyisobutyryl-CoA hydrolase (HIBCH), and branched-chain α-keto acid dehydrogenase (BCKDH) in TAG accumulation have been demonstrated by genetic manipulation (Ge et al., 2014; Pan et al., 2017; Liang et al., 2019). Knockdown of *MCC* or knockout of *BCKDH* led to decreased TAG accumulation (Ge et al., 2014; Liang et al., 2019), and overexpression of *HIBCH* or knockdown of *PCC* increased TAG accumulation (Pan et al., 2017). Carbon skeletons from BCAA degradation enter the TCA cycle through acetyl-CoA in the mitochondria. The single-copy HIBCH in *P. tricornutum* has been proved to be localized in the mitochondria (Pan et al., 2017). Moreover, subcellular localization prediction shows that single-copy β-subunit of BCKDH (BCKDH2), α-subunit of MCC (MCC1), α-subunit of PCC (PCC1), 3-hydroxyisobutyrate dehydrogenase (HIBADH), and methylmalonyl-CoA mutase (MCM) together with two aldehyde dehydrogenases (ALDHs) are located in the mitochondria (**Supplementary Table 1**). Different from *P. tricornutum*, the whole BCAA degradation in plants takes place in both the mitochondria and peroxisome (Zolman et al., 2001).

HISTIDINE, LYSINE, PHENYLALANINE, AND TYROSINE

Histidine (His) is converted to glutamate (Glu) by four enzymatic steps in animals (Litwack, 2018). However, this pathway has not yet been investigated in plants. In *P. tricornutum*, there is a homologous histidine transaminase (HisAT), which could convert His to Glu and imidazol-5-yl-pyruvate using 2-oxoglutarate as a carbon skeleton (**Figure 1A**). The degradation pathway of imidazol-5-yl-pyruvate is not clear yet in *P. tricornutum*. Free His content was very low in *P. tricornutum* cells and was almost undetectable during nitrogen deficiency in our previous studies (**Supplementary Figure 1**; Ge et al., 2014; Pan et al., 2017). The expression of HisAT was down-regulated during nitrogen limitation (**Figure 2A**), which suggests that His might mainly be metabolized to histamine by the action of histidine decarboxylase (HDC) in the mitochondria (**Figure 1A**). It is likely that His degradation has no contribution to TAG accumulation.

Lysine (Lys) is catabolized *via* the α-amino adipic acid pathway identical to that in plants (Hildebrandt et al., 2015). It is converted to saccharopine and subsequently to α-aminoadipate-δ-semialdehyde by a bifunctional enzyme with two functionally independent domains, namely, lysine-ketoglutarate reductase (LKR) and saccharopine dehydrogenase (SDH), using

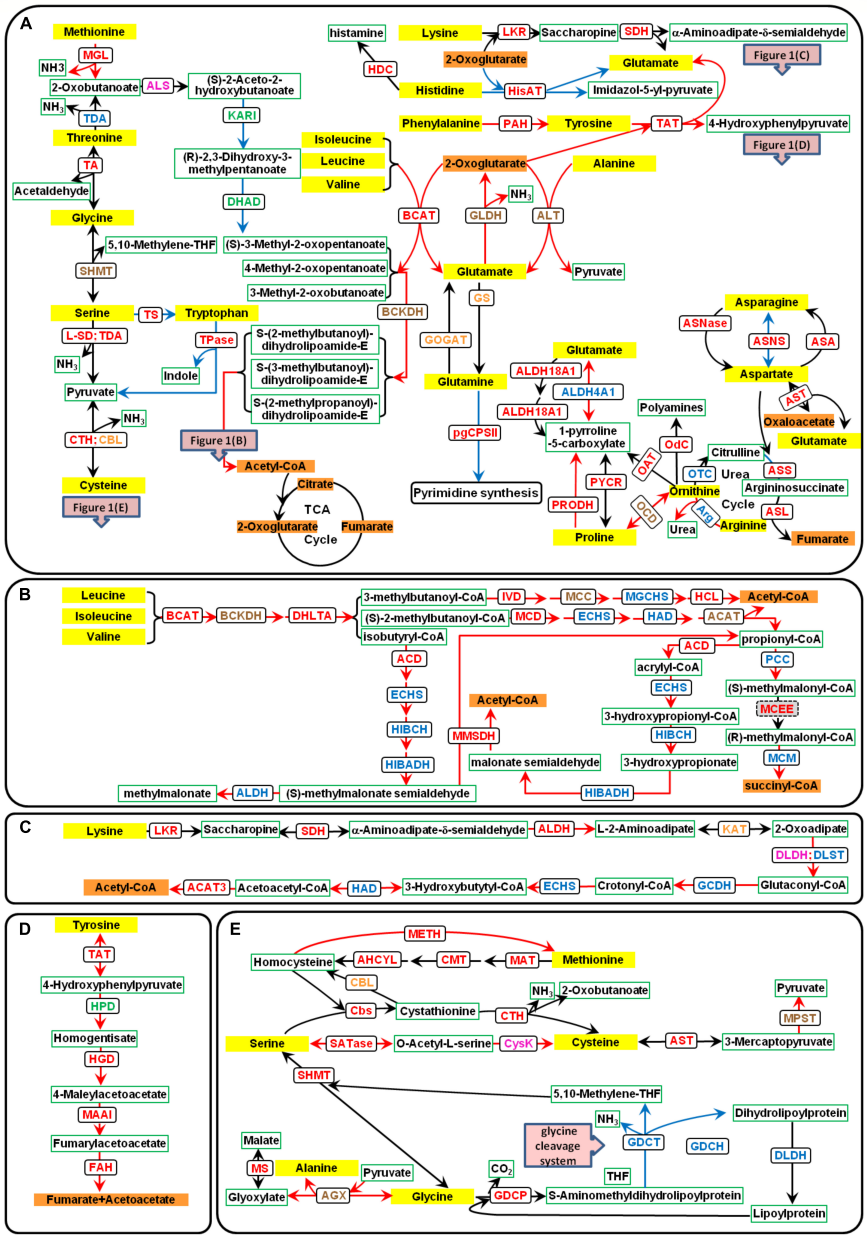


FIGURE 1 | Amino acid catabolic pathways in *P. tricornutum*. The yellow text box represents amino acids, the orange text boxes indicate metabolites that could enter the TCA cycle, the white text boxes with green borders indicate other metabolites, and the white text boxes with black borders indicate the enzymes. The different font colors of enzymes represent different predicted subcellular locations: plastid localization in green font; mitochondria location in blue font; multiple isoenzymes with the plastid and mitochondria localization in orange font; those with mitochondria and other location in brown font; those with plastid and other localization in purple font; and those with plastid, mitochondria, and other localization in red font. Up-regulated genes are indicated with red lines, and down-regulated genes with blue lines. **(A)** Catabolic pathways of all 20 amino acids. **(B)** Catabolic pathways of BCAAs. **(C)** Catabolic pathways of lysine. **(D)** Catabolic pathways of tyrosine. **(E)** Catabolic pathways of serine, methionine, cysteine, and glycine. ACAT, acetyl-CoA *C*-acyltransferase (EC2.3.1.16); ACD, acyl-CoA dehydrogenase (EC1.3.8.1); AGX, alanine–glyoxylate aminotransferase (EC:2.6.1.44); AHCYL, adenosylhomocysteinase (EC3.3.1.1); ALDH, aldehyde dehydrogenase (EC1.2.1.31); ALDH18A1, delta-1-pyrroline-5-carboxylate synthetase (EC2.7.2.11 and EC1.2.1.41); ALDH4A1, 1-pyrroline-5-carboxylate dehydrogenase (EC1.2.1.88); ALS, acetolactate synthase (EC2.2.1.6); ALT, alanine transaminase (EC2.6.1.2); Arg, arginase (EC3.5.3.1); ASA, aspartate–ammonia ligase (EC6.3.1.1); ASL, argininosuccinate lyase (EC4.3.2.1); ASNase, asparaginase (EC3.5.1.1); ASNS, asparagine synthase (EC6.3.5.4); ASS, argininosuccinate synthase (EC6.3.4.5); AST, aspartate aminotransferase (EC2.6.1.1); BCAT, branched-chain amino acid transaminase (EC2.6.1.42); BCKDH, branched-chain α-keto acid dehydrogenase (EC1.2.4.4); CBL, cysteine-*S*-conjugate beta-lyase (EC4.4.1.13); Cbs, cystathionine beta-synthase (EC4.2.1.22); CMT, DNA (cytosine-5)-methyltransferase (EC2.1.1.37); CPSII, carbamoyl-phosphate synthase II (EC6.3.5.5); CTH, cystathionine gamma-lyase (EC4.4.1.1); CysK, cysteine synthase (EC2.5.1.47); DHAD, dihydroxy-acid dehydratase (EC4.2.1.9); DHLTA, dihydrolipoyllysine-residue (2-methylpropanoyl) transferase (EC2.3.1.168); DLDH, dihydrolipoyl dehydrogenase (EC1.8.1.4); DLST, dihydrolipoamide succinyltransferase; ECHS, enoyl-CoA hydratase (EC4.2.1.17); FAH, fumarylacetoacetase

(Continued)

FIGURE 1 | Continued
(EC3.7.1.2); GCDH, glutaryl-CoA dehydrogenase (EC1.3.8.6); GDCH, glycine cleavage system H protein; GDCP, glycine decarboxylase p-protein (EC1.4.4.2); GDCT, glycine decarboxylase t-protein (EC2.1.2.10); GLDH, glutamate dehydrogenase (EC1.4.1.2 and EC1.4.1.4); GOGAT, glutamine 2-oxoglutarate aminotransferase (EC1.4.1.13, EC1.4.1.14, and EC1.4.7.1); GS, glutamine synthetase (EC6.3.1.2); HAD, 3-hydroxyacyl-CoA dehydrogenase (EC1.1.1.35); HCL, hydroxymethylglutaryl-CoA lyase (EC4.1.3.4); HDC, histidine decarboxylase (EC:4.1.1.22); HGD, homogentisate 1,2-dioxygenase (EC1.13.11.5); HIBADH, 3-hydroxyisobutyrate dehydrogenase (EC1.1.1.31); HIBCH, 3-hydroxyisobutyryl-CoA hydrolase (EC3.1.2.4); HisAT, histidine transaminase (EC2.6.1.38); HPD, 4-hydroxyphenylpyruvate dioxygenase (EC1.13.11.27); IVD, isovaleryl-CoA dehydrogenase (EC1.3.8.4); KARI, ketol-acid reductoisomerase (EC1.1.1.86); KAT, kynurenine aminotransferase (EC2.6.1.39); LKR, lysine-2-oxoglutarate reductase (EC1.5.1.8); L-SD, L-serine ammonia-lyase (EC4.3.1.17); MAAI, maleylacetoacetate isomerase (EC5.2.1.2); MAT, *S*-adenosylmethionine synthetase (EC2.5.1.6); MCC, methylcrotonyl-CoA carboxylase (EC6.4.1.4); MCD, 2-methylacyl-CoA dehydrogenase (EC1.3.99.12); MCEE, methylmalonyl-CoA epimerase (EC5.1.99.1); MCM, methylmalonyl-CoA mutase (EC5.4.99.2); METH, methionine synthase (EC2.1.1.13); MGCHS, methylglutaconyl-CoA hydratase (EC4.2.1.18); MGL, methionine gamma-lyase (EC4.4.1.11); MMSDH, methylmalonate semialdehyde dehydrogenase (EC1.2.1.27); MPST, 3-mercaptopyruvate sulfurtransferase (EC2.8.1.2); MS, malate synthase (EC2.3.3.9); OAT, ornithine aminotransferase (EC2.6.1.13); OCD, ornithine cyclodeaminase (EC4.3.1.12); Odc, ornithine decarboxylase (EC4.1.1.17); OTC, ornithine carbamoyltransferase (EC2.1.3.3); PAH, phenylalanine hydroxylase (EC1.14.16.1); PCC, propionyl-CoA carboxylase (EC6.4.1.3); PRODH, proline dehydrogenase (EC1.5.5.2); PYCR, pyrroline-5-carboxylate reductase (EC1.5.1.2); SATase, serine *O*-acetyltransferase (EC2.3.1.30); SDH, saccharopine dehydrogenase (EC1.5.1.9); SHMT, serine hydroxymethyltransferase (EC2.1.2.1); TA, threonine aldolase (EC4.1.2.5); TAT, tyrosine aminotransferase (EC2.6.1.5); TDA, threonine deaminase (EC4.3.1.19); Tpase, tryptophanase (EC4.1.99.1); TS, tryptophan synthase (EC4.2.1.20).

2-oxoglutarate as substrate to produce Glu (**Figure 1C**). The response of the bifunctional LKR/SDH enzyme encoding genes (*Phatr3_EG02101* and *Phatr3_J41409*) to nitrogen limitation was different between nitrogen-starved batch cultures and continuous cultures (**Figure 2A**). In *Arabidopsis*, the expression of LKR/SDH was regulated by abscisic acid (ABA), jasmonate, sugar starvation, and/or nitrogen starvation (Stepansky and Galili, 2003). Some enzymes in the catabolic pathways of Lys are shared by other amino acids, including enoyl-CoA hydratase (ECHS) and 3-hydroxyacyl-CoA dehydrogenase (HAD) involved in the catabolic pathways of isoleucine (Ile), and kynurenine aminotransferase (KAT, also annotated as cysteine-*S*-conjugate beta-lyase, CBL) in cysteine (Cys) metabolism (**Figures 1A,E**). Almost all the genes involved in Lys degradation were up-regulated with the decrease of Lys under nitrogen limitation (**Figure 2A** and **Supplementary Figure 1**). The early steps of Lys degradation (from Lys to α-aminoadipate-d-semialdehyde) may take place in the cytoplasm and the later steps (from α-aminoadipate-d-semialdehyde to acetyl-CoA) in the mitochondria according to subcellular localization prediction (**Supplementary Table 1**). The reaction of 2-oxoadipate oxidatively decarboxylated to glutaryl-CoA is equivalent to the oxidative decarboxylation of 2-oxoglutarate in the TCA cycle. The subsequent degradation reaction was similar to that of Ile in the mitochondria. Since one of the end degradation products of Lys is acetyl-CoA (**Figure 1C**) and the genes involved in Lys degradation are up-regulated, it is likely that Lys catabolism participates in the regulation of carbon/nitrogen partitioning and TAG accumulation.

Like that in animals, phenylalanine (Phe) is hydroxylated to tyrosine (Tyr) by phenylalanine hydroxylase (PAH) prior to degradation in *P. tricornutum* (**Figure 1A**). However, the catabolic pathway remains largely unknown since no PAH homolog has been found in plants (Hildebrandt et al., 2015). The complete degradation pathway of Tyr has been demonstrated in plants based on the degradation pathway in mammals (Dixon and Edwards, 2006). The amino group is transferred to Glu by tyrosine aminotransferase (TAT), and the product was finally degraded into fumarate and acetoacetate by four enzymatic steps. The homologous enzymes for each step have been found in *P. tricornutum* (**Figure 1D**). Phatr3_EG02215 annotated as fumarylacetoacetase has two annotations in UniProt, namely, fumarylacetoacetase and glutathione *S*-transferase (GST). Peptides of the latter annotation of protein Phatr3_EG02215 are homologous with human glutathione *S*-transferase zeta 1 (GSTZ1), which was also described as maleylacetoacetate isomerase (MAAI). The amino acid sequence of GST in *P. tricornutum* shares 48 and 53% identity with mammalian and *T. pseudonana* MAAI, respectively, and all the three sequences contain the conserved motif and active site of MAAI (Polekhina et al., 2001; **Supplementary Figure 2**). This means that Phatr3_EG02215, annotated as one gene (fumarylacetoacetase encoding gene) in *Ensembl Protists*, may actually be a compound of two genes (fumarylacetoacetase encoding gene and *MAAI*). In addition, Phatr3_J36390 was annotated as MAAI in the studies of Levering et al. (2016, 2017), though it is more similar to glutathione *S*-transferase alpha (GSTA) than GSTZ1 by BlastP analysis. The concentration of Phe and Tyr was decreased during nitrogen limitation in *P. tricornutum* (**Supplementary Figure 1**). The expression levels of genes involved in Phe and Tyr degradation were up-regulated accordingly (**Figure 2A**). Phe and Tyr catabolic pathways may take place in the cytoplasm according to subcellular localization prediction (**Supplementary Table 1**). Although fumarate, the final product of Phe and Tyr degradation, is an intermediate of the TCA cycle, the contribution of the catabolic pathways of the two amino acids to TAG accumulation still needs to be investigated.

ALANINE, GLUTAMATE, AND GLUTAMINE

Alanine (Ala) can be directly converted to pyruvate by alanine aminotransferases (ALT). Glutamate dehydrogenase (GLDH) catalyzes oxidative deamination of Glu to produce 2-oxoglutarate (**Figure 1A**). The glutamine synthetase (GS)/glutamine 2-oxoglutarate aminotransferase (GOGAT) pathway is important for ammonium assimilation. GS catalyzes Glu and ammonium to produce glutamine (Gln), and Glu also provides α-amino

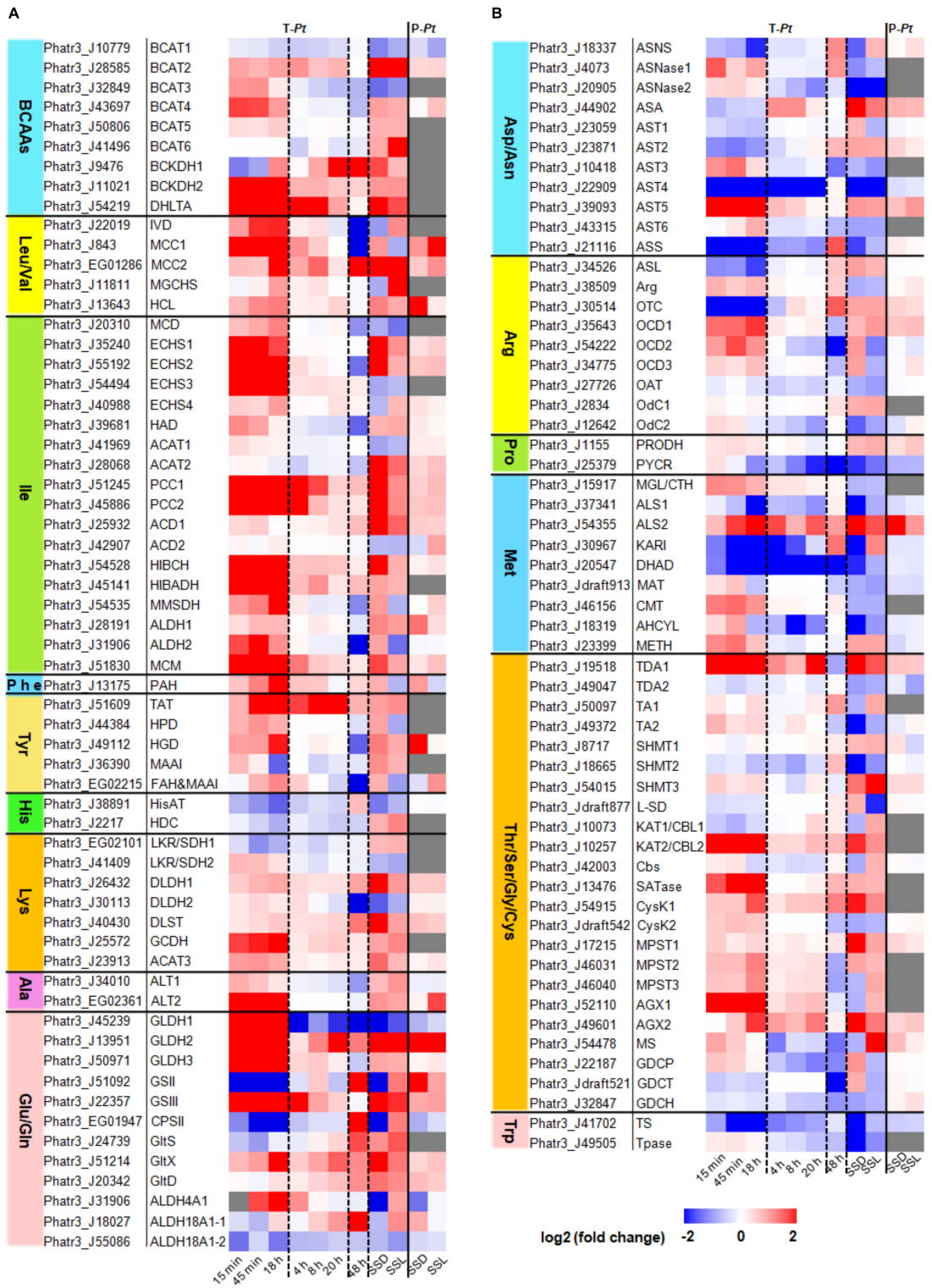


FIGURE 2 | Expression levels of amino acid catabolism-related genes in *P. tricornutum*. Transcriptome data of 15 min, 45 min, and 18 h were cited from Smith et al. (2019). Fold changes of 15 min, 45 min, and 18 h were re-calculated by N-4, N-5, and N-6 contrasting with pre_3, respectively. Transcriptome data of 4, 8, and 20 h were cited from Matthijs et al. (2016, 2017). Transcriptome data of 48 h were cited from Levitan et al. (2015). The transcriptome data and the proteome data of SSL (nitrogen stress to steady state in the light period) and SSD (nitrogen stress to steady state in the dark period) were cited from Remmers et al. (2018). T-*Pt*, transcriptome of *P. tricornutum*; P-*Pt*, proteome of *P. tricornutum*. The homologous genes in *T. pseudonana* and the fold changes and RPKM data were shown in **Supplementary Tables 1, 2**, respectively. All the experimental conditions were shown in the **Supplementary Material**. The abbreviations of related enzymes are the same as those in **Figure 1**.

group for all other amino acid biosynthesis directly or indirectly. Glu and Gln are the most important amino acids as donors for the biosynthesis of major N-containing compounds, including amino acids, nucleotides, chlorophylls, polyamines, and alkaloids (Ireland and Lea, 1999). GS is found, as multiple isoenzyme forms, located both in the cytosol (GS1) and chloroplast/plastid (GS2) and plays distinct roles in most of the higher plants (Ireland and Lea, 1999; Lancien et al., 2000). In *P. tricornutum*, one GS is located in the plastid (GSII, Phatr3_J51092) and the other one in the mitochondria (GSIII, Phatr3_J22357) (Smith et al., 2019). GOGAT transfers the amide-nitrogen of Gln to 2-oxoglutarate, thus providing two molecules of Glu (Forde and Lea, 2007). There are three forms of GOGAT in *P. tricornutum* (Alipanah et al., 2015): one that uses reduced ferredoxin as the electron donor (Fd-GOGAT/GltS, EC 1.4.7.1, Phatr3_J24739), one that uses NADH as the electron donor (NADH-GOGAT/GltX, EC 1.4.1.14, Phatr3_J51214), and the third that uses NADPH as the electron donor (NADPH-GOGAT/GltD, EC 1.4.1.13, Phatr3_J20342). The first two GOGAT present in plants are located in the chloroplast or plastid (Oliveira et al., 1997), and the third is found in bacteria (Reitzer, 1987). The subcellular localization of GOGAT in *P. tricornutum* is not the same to that in plants, which may be contributed to their different origins (Smith, 2018). GltS is located in the plastid, and GltX and GltD are predicated to be in the mitochondria. Thus, the plastidial GSII and GltS are responsible for the assimilation of ammonium produced by nitrate reduction, while mitochondrial GSIII, GltX, and GltD may catalyze the assimilation of Gln from ammonium derived from cytosolic catabolic reactions, e.g., deamination and hydrolysis of organic N (Hockin et al., 2012; Alipanah et al., 2015). pgCPSII uses Gln to perform the first committed step of pyrimidine synthesis (Allen et al., 2011). Besides, Glu can also be converted to 1-pyrroline-5-carboxylate, the intermediate product of proline (Pro) and ornithine (Orn) degradation, by one-step catalytic reaction of 1-pyrroline-5-carboxylate dehydrogenase (ALDH4A1) or by two-step catalytic reaction of delta-1-pyrroline-5-carboxylate synthetase (ALDH18A1) (**Figure 1A**).

The contents of Ala, Glu, and Gln dropped sharply, and Gln was even undetectable during nitrogen limitation in *P. tricornutum* (**Supplementary Figure 1**). The two ALT enzymes and the three GLDH enzymes were up-regulated under nitrogen-starved batch cultures or/and continuous cultures (**Figure 2A**). The mitochondrial GS/GOGAT was up-regulated during nitrogen limitation to assimilate ammonium derived from cytosolic catabolic reactions. The plastidial GS/GOGAT was up-regulated only under nitrogen-starved continuous cultures, suggesting that nitrate reduction was activated in plastid during nitrogen-starved continuous cultures. On the contrary, little ammonium was produced *via* nitrate reduction in plastid under nitrogen-starved batch cultures. Glu may not be used for pyrimidine synthesis during nitrogen limitation as pgCPSII was down-regulated. The mutual conversion of Glu to 1-pyrroline-5-carboxylate was mainly catalyzed by ALDH4A1 in the mitochondria due to the up-regulation of *ALDH4A1* and almost unchanged expression level of *ALDH18A1* according to the transcriptome data. In sum, based on the up-regulated expression of *ALT* gene and the products of metabolism of Ala together with the important roles of Glu and Gln in nitrogen metabolism, it seems that Ala, Glu, and Gln catabolism may contribute to TAG accumulation during nitrogen limitation.

ASPARTATE, ASPARAGINE, ARGININE, AND PROLINE

Aspartate (Asp) and asparagine (Asn) can be converted into each other by asparagine synthase (ASNS), using Gln and Glu as substrates, respectively. Besides, Asp can be converted to Asn by aspartate–ammonia ligase (ASA) with the absorption of ammonia, and Asn can be converted to Asp by asparaginase (ASNase) with the release of ammonia. Asp can be converted to 2-oxoglutarate by aspartate aminotransferase (AST), producing oxaloacetate and Glu (**Figure 1A**). There are six AST encoding genes in *P. tricornutum* that are predicted to be localized in various chambers of cells. Since AST was up-regulated, Asp may be converted to oxaloacetate during nitrogen limitation. The aspartate–argininosuccinate shunt established an association between the ornithine-urea cycle (OUC) and the TCA cycle (Morris, 2002; Allen et al., 2011). In addition, arginine (Arg), Orn, and Pro are directly connected to the OUC, which has been well elaborated in *P. tricornutum* (Allen et al., 2011). Although the contents of Asp, Asn, Arg, Pro, and Orn were decreased during nitrogen limitation (**Supplementary Figure 1**), neither the expression level of genes responsible for the conversion between Asp and Asn nor that of those genes related to OUC was significantly up-regulated (**Figure 2B**). Orn can be metabolized by ornithine carbamoyltransferase (OTC), ornithine decarboxylase (OdC), ornithine aminotransferase (OAT), or ornithine cyclodeaminase (OCD), and only OCD is markedly up-regulated during nitrogen limitation. Arg can be converted to Orn by arginase, which was up-regulated during nitrogen limitation. Therefore, it is very likely that the degradation of Arg produces Pro subsequently converted to 1-pyrroline-5-carboxylate by proline dehydrogenase (PRODH). Then, 1-pyrroline-5-carboxylate is converted to Glu to provide nitrogen in cells for growth. Based on the final degradation products of these four amino acids, their catabolism may contribute little to TAG accumulation.

METHIONINE

The pathway for methionine (Met) degradation that converts Met to 2-oxobutanoate by methionine gamma-lyase (MGL) has been identified in plants but is absent in animals (Rébeillé et al., 2006). 2-Oxobutanoate is a precursor for Ile synthesis, and similar to Ile, it can also be degraded to acetyl-CoA *via* oxidation (**Figure 1A**). An alternative pathway that converts Met to homocysteine, which is subsequently converted to Cys in animals and plants, is catalyzed by three enzymes (**Figure 1E**). Homocysteine can be directly converted to Met, which is catalyzed by methionine synthase (METH). The expression of genes for Met degradation through 2-oxobutanoate was down-regulated, except for one

acetolactate synthase (ALS) encoding gene (**Figure 2B**). ALS catalyzes the degradation of 2-oxobutanoate, an intermediate of threonine (Thr) catabolism. Genes involved in other Met degradation pathways were slightly up-regulated during nitrogen limitation with the decrease of Met (**Supplementary Figure 1**). It is not clear which pathway was dominant for the degradation of Met during nitrogen limitation in *P. tricornutum*.

THREONINE, GLYCINE, SERINE, CYSTEINE, AND TRYPTOPHAN

The catabolic pathways of the five amino acids are very complicated, and some involved enzymes are also present in other amino acid degradation pathways. As mentioned above, Thr can be converted to 2-oxobutanoate by threonine deaminase (TDA). It can also be interconverted with glycine (Gly) by threonine aldolase (TA). Gly can also be interconverted with serine (Ser) by serine hydroxymethyltransferase (SHMT). The easy interconversion of these three amino acids indicates that these reactions can be relevant for the synthesis of the product amino acids and the degradation of the substrate amino acids as well (**Figure 1A**; Hildebrandt et al., 2015).

Cysteine and tryptophan (Trp) can be produced from Ser by two and one enzymatic reactions, respectively (**Figure 1A**). The conversion of the former can be completed through the intermediate product cystathionine or *O*-acetyl-L-serine (**Figure 1E**). Ser, Trp, and Cys can be directly degraded into pyruvate with the release of ammonia or indole. In addition, Cys can also be degraded into pyruvate through a two-step reaction (Hildebrandt et al., 2015; **Figure 1E**). The Gly cleavage system (GCS), which is a mitochondrial multi-enzyme system (also named glycine decarboxylase or glycine dehydrogenase system), comprises four proteins, three enzymes (P-protein, T-protein, and L-protein), and a small lipoylated protein known as H-protein. GCS, an essential and ubiquitous step of both photorespiration and primary metabolism in plants, is responsible for the interconversion of Gly and Ser (Bauwe and Kolukisaoglu, 2003). Although the H-protein has no catalytic activity itself, it acts as a substrate for the P-, T-, and L-proteins and increases the GCS activity (Hasse et al., 2009). Gly can also be transaminated by alanine–glyoxylate aminotransferase (AGX), and the resulting glyoxylate can be acetylated in the peroxisomes to produce malate in plants (Mazelis, 1980). Homologs of genes mentioned above are present in the *P. tricornutum* genome (**Figure 1E**), although the locations of the encoded proteins are not exactly the same as those in plants (**Supplementary Table 1**). Thr and Trp contents decreased and Ser content showed no marked difference, while Gly content increased during nitrogen limitation in *P. tricornutum* (**Supplementary Figure 1**). Cys was not detected in our previous studies (Ge et al., 2014; Pan et al., 2017). In general, Thr and Trp may be degraded mainly through the catalysis of TDA and tryptophanase (Tpase), respectively. The conversion of Ser to Cys and the degradation of Cys may be active during nitrogen limitation according to gene expression levels (**Figure 2B**). However, Ser might not be converted to Trp but was interconverted with Gly frequently. Since GCS-related genes were not up-regulated during nitrogen limitation, Gly was not degraded mainly by GCS. Up-regulated AGX and malate synthase (MS) indicated that Gly was transaminated and acetylated under nitrogen starvation, which may help to explain why the content of Gly did not decrease and even increased during nitrogen limitation. In brief, the contribution of the degradation of these five amino acids to TAG accumulation remains unclear.

CONCLUSION

Amino acids, as protein constituents and essential metabolites, play critical roles in living organisms. Some amino acids (e.g., serine, proline, and leucine) have been shown to act as signaling molecules in plants (Szabados and Savouré, 2010; Häusler et al., 2014; Ros et al., 2014). Therefore, pool sizes of amino acids are of critical importance and are adjusted by amino acid catabolism. The degradation pathways of amino acids in *P. tricornutum* were not identical with those in plants and mammals. In particular, the metabolic pathway of BCAAs in the diatom is different from that in animals, and the subcellular locations of related enzymes are not exactly the same with those in plants. In addition, the metabolic pathways of His and Phe and the OUC in *P. tricornutum* are similar to those in animals, but no related enzymes are found in plants. The mutual transformation pathway of essential amino acids in *P. tricornutum* does not exist in animals. Considering the end degradation products of amino acids and the expression levels of related enzymes in the metabolic pathways during nitrogen limitation, BCAAs, Lys, Ala, Glu, and Gln may contribute to TAG accumulation. Furthermore, to fully understand the catabolic pathways and their regulatory mechanisms, genetic manipulation and a combination of post-genomic approaches (transcriptome, proteome, and metabolome) are necessary for the analyses of mutant and wild-type diatoms.

AUTHOR CONTRIBUTIONS

YP, CY, CL, and TH arranged the data under the guidance of HH. YP and HH prepared and wrote the manuscript. FH reviewed and revised this manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Featured Institute Service Projects from the Institute of Hydrobiology, the Chinese Academy of Sciences (grant number: Y85Z061601) and the National Natural Science Foundation of China (no. 31800302).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.589026/full#supplementary-material

REFERENCES

Admiraal, W., and Peletier, H. (1979). Influence of organic compounds and light limitation on the growth rate of estuarine benthic diatoms. *Br. Phycol. J.* 14, 197–206. doi: 10.1080/00071617900650211

Alipanah, L., Rohloff, J., Winge, P., Bones, A. M., and Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 66, 6281–6296. doi: 10.1093/jxb/erv340

Allen, A. E., Dupont, C. L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473, 203–207. doi: 10.1038/nature10074

Angelovici, R., Lipka, A. E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., et al. (2013). Genome-wide analysis of branched-chain amino acid levels in *Arabidopsis* seeds. *Plant Cell* 25, 4827–4843. doi: 10.1105/tpc.113.119370

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., et al. (2004). The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306, 79–86. doi: 10.1126/science.1101156

Bauwe, H., and Kolukisaoglu, Ü. (2003). Genetic manipulation of glycine decarboxylation. *J. Exp. Bot.* 54, 1523–1535. doi: 10.1093/jxb/erg171

Bender, S. J., Durkin, C. A., Berthiaume, C. T., Morales, R. L., and Armbrust, E. (2014). Transcriptional responses of three model diatoms to nitrate limitation of growth. *Front. Mar. Sci.* 1:3. doi: 10.3389/fmars.2014.00003

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Bromke, M. A. (2013). Amino acid biosynthesis pathways in diatoms. *Metabolites* 3, 294–311. doi: 10.3390/metabo3020294

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). *Phaeodactylum tricornutum*: a diatom cell factory. *Trends Biotechnol.* 38, 606–622. doi: 10.1016/j.tibtech.2019.12.023

Campbell, M. A., Patel, J. K., Meyers, J. L., Myrick, L. C., and Gustin, J. L. (2001). Genes encoding for branched-chain amino acid aminotransferase are differentially expressed in plants. *Plant Physiol. Biochem.* 39, 855–860. doi: 10.1016/S0981-9428(01)01306-7

Contreras, J. A., and Gillard, J. T. (2020). Asparagine-based production of hydrogen peroxide triggers cell death in the diatom *Phaeodactylum tricornutum*. *Bot. Lett.* 1–12. doi: 10.1080/23818107.2020.1754289

Dixon, D. P., and Edwards, R. (2006). Enzymes of tyrosine catabolism in *Arabidopsis thaliana*. *Plant Sci.* 171, 360–366. doi: 10.1016/j.plantsci.2006.04.008

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., and Baart, G. J. (2012). The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner–Doudoroff glycolytic pathway. *Plant J.* 70, 1004–1014. doi: 10.1111/j.1365-313X.2012.04941.x

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. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158

Forde, B. G., and Lea, P. J. (2007). Glutamate in plants: metabolism, regulation, and signalling. *J. Exp. Bot.* 58, 2339–2358. doi: 10.1093/jxb/erm121

Ge, F., Huang, W., Chen, Z., Zhang, C., Xiong, Q., Bowler, C., et al. (2014). Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum*. *Plant Cell* 26, 1681–1697. doi: 10.1105/tpc.114.124982

Geider, R., and La Roche, J. (2002). Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* 37, 1–17. doi: 10.1017/S0967026201003456

Guerra, L. T., Levitan, O., Frada, M. J., Sun, J. S., Falkowski, P. G., and Dismukes, G. C. (2013). Regulatory branch points affecting protein and lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. *Biomass Bioenerg.* 59, 306–315. doi: 10.1016/j.biombioe.2013.10.007

Hasse, D., Mikkat, S., Hagemann, M., and Bauwe, H. (2009). Alternative splicing produces an H-protein with better substrate properties for the P-protein of glycine decarboxylase. *FEBS J.* 276, 6985–6991. doi: 10.1111/j.1742-4658.2009.07406.x

Häusler, R. E., Ludewig, F., and Krueger, S. (2014). Amino acids – A life between metabolism and signaling. *Plant Sci.* 229, 225–237. doi: 10.1016/j.plantsci.2014.09.011

Hildebrandt, T. M., Nunes Nesi, A., Araújo, W. L., and Braun, H. P. (2015). Amino acid catabolism in plants. *Mol. Plant* 8, 1563–1579. doi: 10.1016/j.molp.2015.09.005

Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 158, 299–312. doi: 10.1104/pp.111.184333

Ireland, R. J., and Lea, P. J. (1999). "The enzymes of glutamine, glutamate, asparagine, and aspartate metabolism," in *Plant Amino Acids. Biochemistry and Biotechnology*, ed. B. K. Singh (New York, NY: Marcel Dekker), 49–109.

Jauffrais, T., Jesus, B., Méléder, V., Turpin, V., Arnaldo, D., Raimbault, P., et al. (2016). Physiological and photophysiological responses of the benthic diatom *Entomoneis paludosa* (Bacillariophyceae) to dissolved inorganic and organic nitrogen in culture. *Mar. Biol.* 163:115. doi: 10.1007/s00227-016-2888-9

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925. doi: 10.1038/ncomms7925

Kudela, R. M., and Dugdale, R. C. (2000). Nutrient regulation of phytoplankton productivity in Monterey Bay, California. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 47, 1023–1053. doi: 10.1016/S0967-0645(99)00135-6

Lancien, M., Gadal, P., and Hodges, M. (2000). Enzyme redundancy and the importance of 2-oxoglutarate in higher plant ammonium assimilation. *Plant Physiol.* 123, 817–824. doi: 10.1104/pp.123.3.817

Levering, J., Broddrick, J., Dupont, C. L., Peers, G., Beeri, K., Mayers, J., et al. (2016). Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One* 11:e0155038. doi: 10.1371/journal.pone.0155038

Levering, J., Dupont, C. L., Allen, A. E., Palsson, B. O., and Zengler, K. (2017). Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO2 levels. *mSystems* 2:e00142-16. doi: 10.1128/mSystems.00142-16

Levitan, O., Dinamarca, J., Zelzion, E., Lun, D. S., Guerra, L. T., Kim, M. K., et al. (2015). Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *Proc. Natl. Acad. Sci. U.S.A.* 112, 412–417. doi: 10.1073/pnas.1419818112

Liang, Y., Kong, F., Torres-Romero, I., Burlacot, A., Cuine, S., Legeret, B., et al. (2019). Branched-chain amino acid catabolism impacts triacylglycerol homeostasis in *Chlamydomonas reinhardtii*. *Plant Physiol.* 179, 1502–1514. doi: 10.1104/pp.18.01584

Litwack, G. (ed.) (2018). "Metabolism of amino acids," in *Human Biochemistry*, (Boston, MA: Academic Press), 359–394.

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 Physiol.* 170, 489–498. doi: 10.1104/pp.15.01300

Matthijs, M., Fabris, M., Obata, T., Foubert, I., Franco-Zorrilla, J. M., Solano, R., et al. (2017). The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*. *EMBO J.* 36, 1559–1576. doi: 10.15252/embj.201696392

Mazelis, M. (1980). "15 - Amino acid catabolism," in *Amino Acids and Derivatives*, ed. B. J. Miflin (Cambridge, MA: Academic Press), 541–567. doi: 10.1016/b978-0-12-675405-6.50021-8

Medlin, L., Kooistra, W., and Schmid, A. M. (2000). "A review of the evolution of the diatoms-a total approach using molecules, morphology and geology," in *The Origin and Early Evolution of the Diatoms: Fossil, Molecular and Biogeographical Approaches*, eds A. Witkowski and J. Sieminska (Krakow: Polish Academy of Sciences), 13–35.

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., et al. (2013). Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* 6, 701–710. doi: 10.1038/ngeo1765

Morris, S. M. (2002). Regulation of enzymes of the urea cycle and arginine metabolism. *Annu. Rev. Nutr.* 22, 87–105. doi: 10.1146/annurev.nutr.22.110801.140547

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324, 1724–1726. doi: 10.1126/science.1172983

Oliveira, I. C., Lam, H. M., Coschigano, K., Melo-Oliveira, R., and Coruzzi, G. (1997). Molecular-genetic dissection of ammonium assimilation in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 35, 185–198.

Palenik, B., and Morel, F. M. (1990). Comparison of cell-surface L-amino acid oxidases from several marine phytoplankton. *Mar. Ecol. Prog. Ser.* 59, 195–201. doi: 10.3354/meps059195

Pan, Y., Yang, J., Gong, Y., Li, X., and Hu, H. (2017). 3-Hydroxyisobutyryl-CoA hydrolase involved in isoleucine catabolism regulates triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160409. doi: 10.1098/rstb.2016.0409

Polekhina, G., Board, P. G., Blackburn, A. C., and Parker, M. W. (2001). Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity. *Biochemistry* 40, 1567–1576. doi: 10.1021/bi002249z

Rébeillé, F., Jabrin, S., Bligny, R., Loizeau, K., Gambonnet, B., Van Wilder, V., et al. (2006). Methionine catabolism in *Arabidopsis* cells is initiated by a γ-cleavage process and leads to S-methylcysteine and isoleucine syntheses. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15687–15692. doi: 10.1073/pnas.0606195103

Rees, T. A. V., and Allison, V. J. (2006). Evidence for an extracellular L–amino acid oxidase in nitrogen-deprived *Phaeodactylum tricornutum* (Bacillariophyceae) and inhibition of enzyme activity by dissolved inorganic carbon. *Phycologia* 45, 337–342. doi: 10.2216/04-92.1

Reitzer, L. J. (1987). "Ammonia assimilation and the biosynthesis of glutamine, glutamate, aspartate, asparagine, L-alanine and D-alanine," in *Escherichia Coli and Salmonella typhimurium. Cellular and Molecular Biology,* Vol. 2, eds F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter and H. E. Umbarger (Washington, DC: American Society for Microbiology), 302–320.

Remmers, I. M., D'Adamo, S., Martens, D. E., de Vos, R. C., Mumm, R., America, A. H., et al. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Res.* 35, 33–49. doi: 10.1016/j.algal.2018.08.012

Ros, R., Muñoz-Bertomeu, J., and Krueger, S. (2014). Serine in plants: biosynthesis, metabolism, and functions. *Trends Plant Sci.* 19, 564–569. doi: 10.1016/j.tplants.2014.06.003

Singh, D., Carlson, R., Fell, D., and Poolman, M. (2015). Modelling metabolism of the diatom *Phaeodactylum tricornutum*. *Biochem. Soc. Trans.* 43, 1182–1186. doi: 10.1042/BST20150152

Sipler, R. E., and Bronk, D. A. (2015). "Dynamics of dissolved organic nitrogen," in *Biogeochemistry of Marine Dissolved Organic Matter*, ed. D. A. H. A. Carlson (Boston, MA: Academic Press), 127–232. doi: 10.1016/B978-0-12-405940-5.00004-2

Smith, S. (2018). *Figshare: Nitrogen Gene Phylogenies*. Available online at: https://doi.org/10.6084/m9.figshare.6233198 (accessed October 15, 2020).

Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Oborník, M., Horák, A., et al. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* 10:4552. doi: 10.1038/s41467-019-12407-y

Stepansky, A., and Galili, G. (2003). Synthesis of the *Arabidopsis* bifunctional lysine-ketoglutarate reductase/saccharopine dehydrogenase enzyme of lysine catabolism is concertedly regulated by metabolic and stress-associated signals. *Plant Physiol.* 133, 1407–1415. doi: 10.1104/pp.103.026294

Szabados, L., and Savouré, A. (2010). Proline: a multifunctional amino acid. *Trends Plant Sci.* 15, 89–97. doi: 10.1016/j.tplants.2009.11.009

Zhang, C., and Hu, H. (2014). High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar. Genomics* 16, 63–66. doi: 10.1016/j.margen.2013.10.003

Zolman, B. K., Monroe-Augustus, M., Thompson, B., Hawes, J. W., Krukenberg, K. A., Matsuda, S. P., et al. (2001). chy1, an Arabidopsis mutant with impaired β-oxidation, is defective in a peroxisomal β-hydroxyisobutyryl-CoA hydrolase. *J. Biol. Chem.* 276, 31037–31046. doi: 10.1074/jbc.M104679200

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pan, Hu, Yu, Li, Huang and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 March 2021
doi: 10.3389/fpls.2021.629045

Retracing Storage Polysaccharide Evolution in Stramenopila

Malika Chabi[1†], Marie Leleu[1,2†], Léa Fermont[1], Matthieu Colpaert[1], Christophe Colleoni[1], Steven G. Ball[1] and Ugo Cenci[1]*

[1] Univ. Lille, CNRS, UMR 8576—UGSF—Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France,
[2] InBioS-PhytoSYSTEMS, Eukaryotic Phylogenomics, University of Liège, Liège, Belgium

Eukaryotes most often synthesize storage polysaccharides in the cytosol or vacuoles in the form of either alpha (glycogen/starch)- or beta-glucosidic (chrysolaminarins and paramylon) linked glucan polymers. In both cases, the glucose can be packed either in water-soluble (glycogen and chrysolaminarins) or solid crystalline (starch and paramylon) forms with different impacts, respectively, on the osmotic pressure, the glucose accessibility, and the amounts stored. Glycogen or starch accumulation appears universal in all free-living unikonts (metazoa, fungi, amoebozoa, etc.), as well as Archaeplastida and alveolata, while other lineages offer a more complex picture featuring both alpha- and beta-glucan accumulators. We now infer the distribution of these polymers in stramenopiles through the bioinformatic detection of their suspected metabolic pathways. Detailed phylogenetic analysis of key enzymes of these pathways correlated to the phylogeny of Stramenopila enables us to retrace the evolution of storage polysaccharide metabolism in this diverse group of organisms. The possible ancestral nature of glycogen metabolism in eukaryotes and the underlying source of its replacement by beta-glucans are discussed.

Keywords: laminarin, glycogen, stramenopila, metabolism, CAZy, polysaccharide

OPEN ACCESS

Edited by:
Benoit Schoefs,
Le Mans Université, France

Reviewed by:
Eiji Suzuki,
Akita Prefectural University, Japan
Ansgar Gruber,
Academy of Sciences of the Czech Republic (ASCR), Czechia

****Correspondence:***
Ugo Cenci
ugo.cenci@univ-lille.fr

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *13 November 2020*
Accepted: *18 January 2021*
Published: *03 March 2021*

Citation:
Chabi M, Leleu M, Fermont L, Colpaert M, Colleoni C, Ball SG and Cenci U (2021) Retracing Storage Polysaccharide Evolution in Stramenopila. Front. Plant Sci. 12:629045. doi: 10.3389/fpls.2021.629045

INTRODUCTION

Storage polysaccharides are central components of the cell, used to store energy and carbon in an osmotically inert form. They can be found under diverse forms including mannans and glucans; however, by far the two most widespread kinds consist of glucose polymers (glucans) linked by either beta (subsequently called beta-glucan polysaccharides)- or alpha (subsequently called alpha-glucan polysaccharides)-glucosidic bonds. In both cases, storage polysaccharides can be found under either water-soluble polymers (e.g., glycogen and chrysolaminarins) or insoluble solid crystalline material (e.g., starch and paramylon). This difference in physical state of these polymers directly impacts both glucose accessibility (which is slow with solid crystalline granules) and stored amounts (which appear generally higher for crystalline material) as well as the osmotic activity of the cellular compartments concerned by their accumulation (cytosol, vacuoles, and plastid).

All of these storage polysaccharides are obtained through similar kinds of "core" pathways grouped under the term carbohydrate active enzymes (cazymes), which are referenced and classified in families by the CAZy database (Lombard et al., 2014). The glucose is in all cases first activated in the form of a nucleotide sugar (UDP-glucose or ADP-glucose), which is then transferred to a growing chain through glycosyl transferases (GT) behaving as glucan synthases.

These consist of GT3 or GT5 enzymes for alpha-glucans (Ball et al., 2011) and GT48 for beta-glucans (Huang et al., 2018). Those synthases will create a link between carbon 1 and 4 and carbon 1 and 3, for alpha- and beta-glucan, respectively. Subsequently, these chains can be branched on position 6 of some glucose residues within the otherwise linear chains. In the case of alpha-glucans, this is usually performed by GH13 subfamilies 8 and 9 (written GH13_8 or GH13_9) glycosyl hydrolases (GH). However, while GH16 activities are thought to be involved in beta-glucan branching (Huang et al., 2016), functional proof of their involvement is still lacking.

Intracellular catabolism of storage polysaccharides is performed either through release of glucose monomers or oligomers (by various hydrolytic activities for both alpha- and beta-glucans), or by phosphorolysis that will produce Glucose-1-phosphate (Glc-1-P), which, unlike glucose monomers, retains one of the two high-energy bonds used during glucan synthesis. In alpha-glucan metabolism, this is performed by GT35 (Wilson et al., 2010) phosphorylases, while for beta-glucan degradation, GH94, GH149, and GH161 have all been reported to perform phosphorolysis (Kitaoka et al., 2012; Kuhaudomlarp et al., 2018, 2019) *in vitro*. Again, their *in vivo* functional involvement still requires demonstration. To complete degradation, enzymes performing the reverse reaction of branching called debranching enzymes are needed. For alpha-glucan degradation, two distinct types of enzymes are reported. These include both so-called "direct" and indirect debranching enzyme. Direct debranching enzymes of GH13 subfamily 11 are widespread enzymes of bacterial glycogen degradation while their distribution remains restricted to Archaeplastida within the eukaryotic domain where they play a role in both synthesis and degradation of starch (Mouille et al., 1996; Cenci et al., 2014). Such enzymes "directly" access the α-1,6 branch and hydrolyze it, thereby releasing oligosaccharides in the cytosol. Indirect debranching enzyme (iDBE) defines a bifunctional enzyme selectively active for eukaryotic glycogen degradation. It includes two domains, a GH13_25 associated with a GH133 domain (Teste et al., 2000). Catalysis proceeds first by transfer of chain, segments preceding the branch, followed by the hydrolytic cleavage of the unique glucose residue left at the branch thereby obviating oligosaccharide release. However, no activity has been proposed, yet, for debranching in the case of beta-glucans.

While all steps of the core pathway of alpha-glucan metabolism have received abundant functional demonstration of their involvement in many distinct lineages, functional evidence supporting the core pathway of beta-glucan metabolism remains scarce and mostly limited to the glucan elongation step (Huang et al., 2016, 2018). Caution is thus needed as many of the core pathway inferred steps still require functional demonstrations.

The much wider distribution of the alpha-glucan metabolism pathways in the eukaryotic tree of life [all unikonts (Amoebozoa, Obazoa) and many bikonts (SAR, Cryptista, Haptista, Archaeplastida, Excavata)] when compared to that of beta-glucans (a few bikonts) may suggest that alpha-glucan metabolism defines a comparatively more ancient eukaryotic pathway.

In addition, it is worth noting that alpha- and beta-glucans are also found in cell walls. While cell walls with beta-1,3 glucan polysaccharides are quite common, cell wall containing alpha-1,4 glucans are quite rare being presently restricted to mycobacterium (Dinadayala et al., 2008) and a very small number of fungi (Yoshimi et al., 2017).

However, the more the diversity is explored, the more complex the picture gets. Among others, Stramenopila have been known chiefly as a beta-glucan storage polysaccharide accumulating organism, but recent analysis has shown additional layers of complexity in some lineages (presence of alpha-glucan storage metabolic enzymes in Blastocystis) (Gentekaki et al., 2017).

Stramenopila are a very diverse group of eukaryotes including both photosynthetic and free-living phagotrophic or non-phagotrophic or parasitic heterotrophic lineages. In all cases, unicellular species such as diatoms or intestinal Blastocystis protists as well as more or less complex multicellular clades such as brown algae and oomycete pathogens have been reported. Several hypotheses have been proposed to explain the evolution of Stramenopila, including a unique secondary red alga endosymbiosis (Chromalveolate hypothesis) (Cavalier-Smith, 1999) or a succession of endosymbioses where Haptophyta would have been internalized by a Stramenopila non-photosynthetic protist ancestor (Serial hypothesis) (Baurain et al., 2010). Non-photosynthetic Stramenopila with no evidence for a prior endosymbiotic history have been identified including intestinal protists (Blastocystis) or multicellular parasites (e.g., Oomycetes) (Stiller et al., 2014). Stramenopila can be separated into two clades: (i) Bigyra subsequently composed of Opalozoa (e.g., Blastocystis and Halocafeteria) and Sagenista (Aurantiochytrium and Aplanochytrium), and (ii) Gyrista composed of Oomycota (e.g., Phytophtora) and Ochrophyta (e.g., the multicellular brown alga Ectocarpus and the unicellular diatom Phaeodactylum) (Derelle et al., 2016; **Figure 1**). The Ochrophyta are presently believed to account for up to 25% of Earth's total annual carbon fixation chiefly through the metabolic activity of diatoms (Field, 1998). In diatoms, this carbon fixation is known to occur through storage in vacuoles of glucose polymers in the form of chrysolaminarin polysaccharides consisting of beta-1,3-glucans with occasional beta-1,6-linked branches. In other Stramenopila from the Gyrista group, storage polysaccharides of very similar structures have been named mycolaminarin and laminarin in Oomycota (Kamoun, 2003) and brown algae (Graiff et al., 2016), respectively. However, in Opalozoa, despite the fact that very little is known on the nature of their carbon storage, glycogen accumulation has been reported in Blastocystis intestinal protists. As to Sagenista, current literature survey points to a major accumulation of storage lipids, as some species are described for their high quantity of polyunsaturated fatty acid (Lee Chang et al., 2012). However, storage lipid accumulation is often found in Stramenopila, including groups that are known to accumulate storage polysaccharides. No clear cytological or genomic analysis has really, to our knowledge, questioned the possible presence of storage polysaccharide metabolism in Sagenista. In addition, in Opalozoa, despite the recent preliminary report of alpha-glucan metabolism in Blastocystis (Gentekaki et al., 2017), the paucity of studies dealing with such

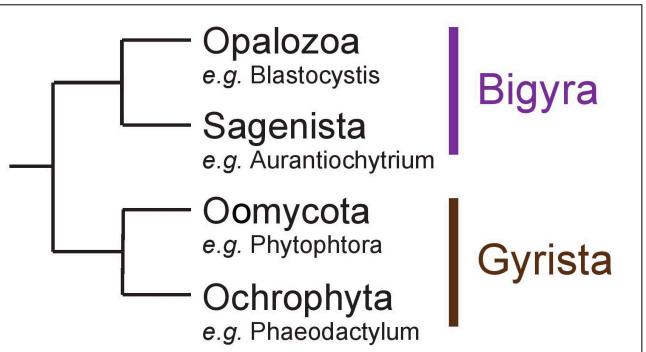


FIGURE 1 | Schematic representation of the Stramenopila evolution tree.

issues prevent us from fully understanding carbon storage in these organisms.

We now infer the distribution of these polymers in Stramenopila through the bioinformatic detection of their suspected metabolic pathways and confirm the presence of alpha-glucan metabolism in Opalozoa other than Blastocystis, as well as the apparent absence of both alpha and beta storage polysaccharide metabolism in all sequenced Sagenista genomes. We further analyze the origin of beta-glucan metabolism in Gyrista and propose that the Stramenopiles ancestor would have been able to synthesize both alpha- and beta-glucan storage polysaccharides.

MATERIALS AND METHODS

Sequence Selection

Sequences were selected from a selection of Stramenopiles selected for their completeness and their low contamination used from Vlierberghe et al. (2021), as well as from other sources to ensure a representative diversity of Stramenopiles (Corteggiani Carpinelli et al., 2014; Harding et al., 2016). In addition, we used the advance access genome drafts for Ochromonas (https://genome.jgi.doe.gov/Ochro2298_1/Ochro2298_1). Genomes selected are as follows: *Albugo candida, Aplanochytrium kerguelense, Aurantiochytrium limacinum, Aureococcus anophagefferens, Blastocystis* sp., *Ectocarpus siliculosus Fragilariopsis cylindrus, Nannochloropsis gaditana, Ochromonadaceae* sp. CCMP2298, *Halocafeteria seosinensis, Phaeodactylum tricornutum Phytophthora infestans, Pseudo-nitzschia multiseries, Pythium ultimum*, *Saprolegnia parasitica, Schizochytrium aggregatum*, and *Thalassiosira pseudonana*. Every proteome was then screened using dbCAN hmm model (Yin et al., 2012), which is based on the CAZy database (Lombard et al., 2014) using HMMER (Mistry et al., 2013), for specific models suspected of being involved in either alpha storage polysaccharide and beta storage polysaccharide. In addition, we have selected sequences from the CAZy database to design our own profiles for each GH16 subfamily using MUSCLE software (Edgar, 2004) to align sequences and hmmbuild to construct profiles (Mistry et al., 2013). We then used a threshold of 10e-40 to annotate the sequence automatically; in addition, we consider classification with subfamilies only if the second best-matching HMM had an E-value more than 10e-10-fold greater. Then, we manually look for annotation between 1e-40 and 10e-20, as well as check subfamilies with less than 10e-10-fold greater.

Deciphering New Putative Enzymes Involved in (Chryso-myco)laminarin

To decipher new putative enzymes involved in (chryso-myco)laminarin biosynthesis in Stramenopiles, we used the literature on beta-glucan metabolism to select a wide range of CAZy proteins putatively involved in the beta storage polysaccharide metabolism in Stramenopiles. Using the annotation process described above, we then compared their distribution to GT48 and GH16_2 (the only two enzymes known to be involved in laminarin metabolism). If the distribution was correlating with GT48 and GH16_2, we considered it as new putative enzymes. As criteria for correlation, we considered a CAZy family (or subfamily) if (i) no organisms without GT48 and GH16_2 would have this group of enzymes, and (ii) at most two genomes, from the known beta-glucan storage polysaccharide accumulator, lack a family.

Subcellular Localization Prediction

In order to predict the subcellular locations of beta-glucan metabolism enzymes, we first selected sequences from our annotation survey (**Supplementary Table S1**) for GT48, GH16_2 and GH5_33, GH81, and GH161. Then, we carried out different predictions using a combination of Almagro Armenteros et al., 2019; Petsalaki et al., 2006, and TMHMMv2.0 (Krogh et al., 2001) under the "non-plant" modes.

Phylogenetic Analysis

For the CAZy families considered as key enzymes for storage polysaccharide metabolism (alpha-glucan storage polysaccharide metabolism enzymes: GH13_25-GH133, GT35, GT5, and GH13_8; beta-glucan storage polysaccharide metabolism enzymes characterized: GH16_2, GT48, and the putative enzymes associated with beta-glucan storage polysaccharide GH5_33, GH81, and GH161), we have performed phylogeny using a pipeline used in previous studies (Cenci et al., 2016, 2018a). We used all sequences, annotated as mentioned above, and formed a cluster for each different CAZy family. Each cluster was used to retrieve sequences using homology searches by BLAST against sequences of the non-redundant protein sequence database of the NCBI and sequences from other databases (MMETSP and data publicly available). We retrieve the top 2,000 homologs with an E-value cutoff lower than 1e-10 and aligned them using MAFFT with the quick alignments settings (Katoh and Standley, 2013). Block selection was then performed using BMGE (Criscuolo and Gribaldo, 2010) with a block size of 4 and the BLOSUM30 similarity matrix. We generated preliminary trees using Fasttree (Price et al., 2010), and "dereplication" was applied to robustly supported monophyletic clades using TreeTrimmer (Maruyama et al., 2013) in order to reduce sequence redundancy. The final set of sequences was manually selected and focused around Stramenopila sequences. Finally, proteins were re-aligned with MUSCLE, block selection

was carried out using BMGE with the same settings as above, and trees were generated with Phylobayes-4.1 under the catfix C20 + Poisson model with the two chains stopped when convergence was reached (maxdiff <0.1) after at least 500 cycles, discarding 500 burn-in trees. Bootstrap support values were estimated from 100 replicates using IQ-TREE under the LG4X model and mapped onto the Bayesian tree.

RESULTS

Detection of Alpha-Glucan Metabolism in Opalozoa

In a previous study of the Blastocystis cazyme content, we have uncovered several genes known in other eukaryotes to be involved in alpha-glucan storage polysaccharide metabolism (Gentekaki et al., 2017). However, the scarcity of genomes in Opalozoa and the evolutionary trajectory of Blastocystis compared to other Stramenopila [a history marked by high-speed diversification and LGT (Eme et al., 2017)] prevented us from concluding the ancient nature of alpha-glucan metabolism in Stramenopila. Thus, we added a recent transcriptome of another Opalozoa: *H. seosinensis*, to the analysis and found the presence of all required alpha-glucan metabolism enzymes in these organisms (**Table 1**). We, thus, found a full set of enzymes able to perform glycogen biosynthesis including a GT5 glycogen synthase, a branching enzyme (GH13 subfamily 8 written: GH13_8), and an indirect debranching enzymes with the two expected domains (GH13_25-GH133) as well as glycogen phosphorylase (GT35), all of which are present in both Blastocystis and Halocafeteria. However, those enzymes were never found in other Stramenopila from either the Gyrista or the Sagenista sister group, suggesting the absence of alpha-glucan storage polysaccharide metabolism in these taxa.

Identification of Strong Candidate Enzymes Involved in (Chryso-myco)laminarin Biosynthesis in Stramenopila

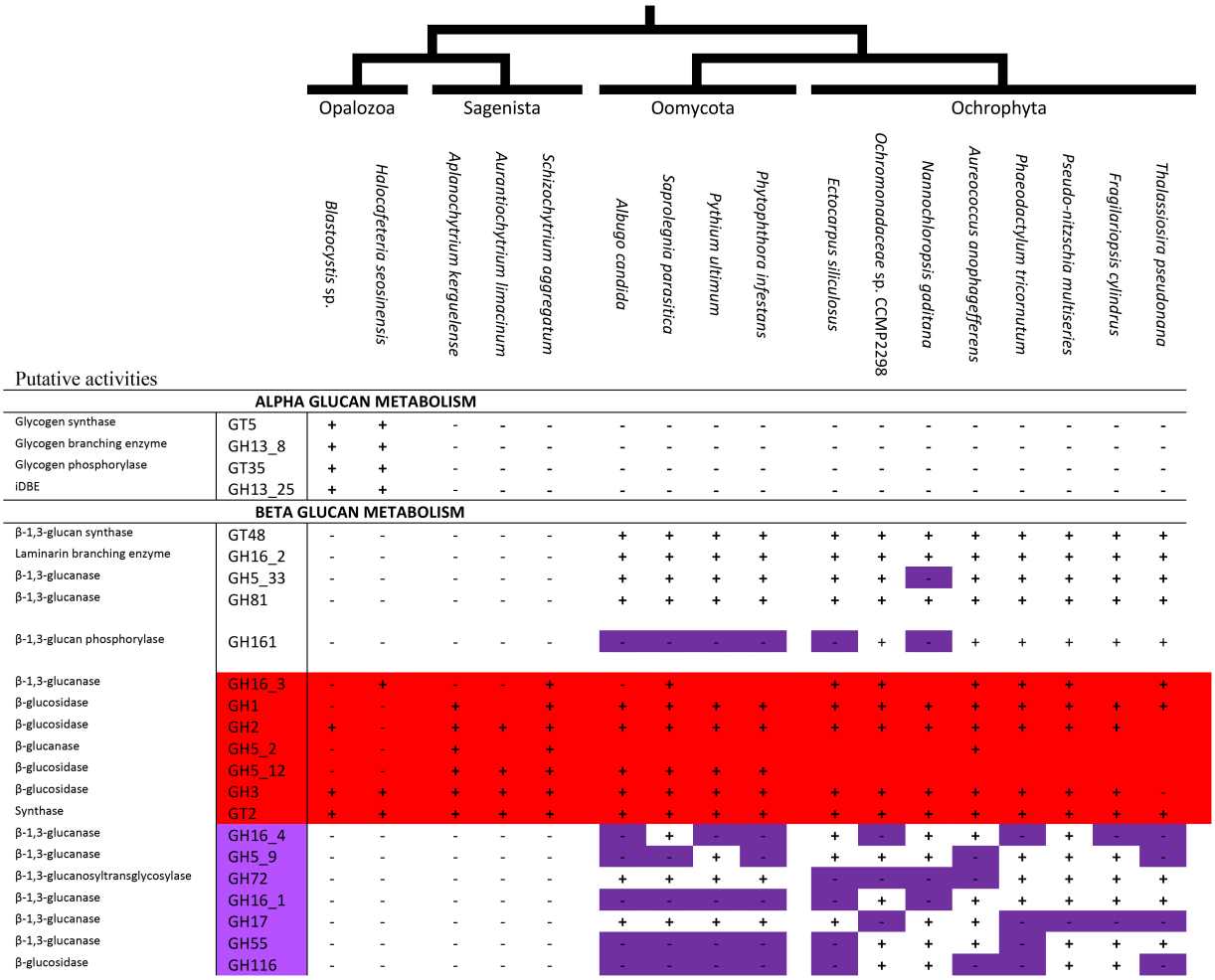
Alpha-glucan metabolism being limited to Opalozoa, we further checked for the presence of candidate enzymes possibly involved in beta-glucan metabolism in Stramenopila. The GT48 characterized in *P. tricornutum* (Huang et al., 2018) has been functionally demonstrated to be active for synthesis of linear beta-1,4 glucans while a GH16 is presently highly suspected but not fully demonstrated to be active for synthesis of the beta-1,6 branches. We can thus correlate the presence of beta-glucan metabolism to the presence of the GT48. In addition, we built manually HMM profiles for each of the GH16, thanks to the public CAZy database and their recent work on GH16 (Viborg et al., 2019). Thus, we were able to assign the enzymes suspected of being responsible for polysaccharide branching in Stramenopila to the GH16 subfamily 2. In addition, we checked a high number of putative beta-glucan candidate metabolic enzymes proposed in the literature for beta-glucan polysaccharide storage metabolism (Michel et al., 2010; O'Neill et al., 2015; Kuhaudomlarp et al., 2019) and tried to correlate them with the presence of the GT48 and GH16 subfamily 2 (**Table 1** and **Supplementary Table S1**). Astonishingly, we did not find a correlation (absence in several Gyrista) with the GH16 subfamily 3 and 4 sequences despite these having been reported to encode laminarinase activities (Viborg et al., 2019). We specifically found instead two different candidate enzymes correlating with the distribution of the GT48 in Stramenopila: GH81 and GH5_33. These had both been previously reported as possibly involved in laminarin hydrolysis despite the absence of clear endomembrane targeting sequences (Michel et al., 2010; O'Neill et al., 2015). In addition, in agreement with recent proposals (Kuhaudomlarp et al., 2019), we consider that the GH161 phosphorylase, despite its absence from oomycetes genomes, remains possibly involved in phosphorolysis of beta-glucans albeit selectively in Ochrophytes, mirroring the role of the GT35 glycogen/starch phosphorylase in these beta-glucan accumulating lineages. Nevertheless, if we assume this function for ochrophyte polysaccharide breakdown, we must conclude that at variance with what has been demonstrated for alpha-glucan metabolism, phosphorolysis remains dispensable for beta-glucan metabolism as exemplified by oomycetes since no other candidate beta-glucan phosphorylases from the family GH94 and GH149 has been revealed in Stramenopila.

In addition, we have analyzed the subcellular localization and presence of transmembrane domains, on the subset of activities suggested in our study to be involved in beta-glucan metabolism (GH81 and GH5_33), and compared the analysis to results obtained with activities previously established as involved (GT48, GH16_2, and GH161) (**Supplementary Table S2**). We observed that prediction of the secretion signal, expected to deliver proteins to vacuoles (Gruber and Kroth, 2017), was not conserved among the different sequences and was not informative enough to help rule out their vacuolar localization. For instance, GT48 sequences were mainly lacking secretion signals. However, analysis of transmembrane domains shows their clear presence in GT48 enzymes, which strongly suggest their presence in the tonoplast membrane. For all other activities (GH81, GH5_33, GH16_2, and GH161), because only a few transmembrane domains were detected and that most of the sequences did not display a clear targeting signal, those results prevent us from drawing clear conclusions.

Alpha-Glucan Metabolism in Stramenopila Is From an Ancient Eukaryotic Origin

Based on the presence of alpha-glucan metabolism in both Opalozoa genomes analyzed, we tried to understand if enzymes were either acquired in a common ancestor or separately in those organisms. We, thus, performed a phylogenetic analysis of alpha-glucan storage metabolism enzymes. Using the MMETSP database (Keeling et al., 2014), we were able to find some enzyme sequences from the Cafeteria and Bicosoecida genera, to strengthen our analysis. We observed two patterns: on the one hand, for two enzymes, indirect debranching enzyme (GH13_25-GH133) (**Supplementary Figure S1**) and branching enzyme

TABLE 1 | Table of CAZy families (indicated on the first column) and subfamilies found in 17 Stramenopila genomes.



Presence of those enzymes in the genome is indicated by a +, while absence is indicated by a - and a purple color. To better observe correlation for beta-glucan storage metabolism, we also added a red background to highlight enzymes not specifically linked to beta-glucan metabolism, because they are present in organisms without beta-glucan storage polysaccharides, even if some isoforms can probably act on it. In addition, we display in light purple activities that were missing in clade normally synthesizing beta-glucan storage polysaccharides. For CAZy families and subfamilies, we indicate with a white background if they were linked to one of the metabolisms. Despite a low correlation for GH161, we kept it because of its putative involvement as stated in Kuhaudomlarp et al. (2019). Finally, we have indicated on the first column putative activities of each cazyme.

(GH13_8) (**Figure 2**), we were able to find Blastocystis and Halocafeteria or Cafeteria grouping together. However, in every case, bootstraps (BS) and posterior probabilities (pp) were low. In addition, in these two trees, at least one sequence of Stramenopila was not found inside the corresponding monophyletic group. On the other hand, in the case of both GT35 (**Figure 3**) and GT5 (**Supplementary Figure S2**), Blastocystis was not convincingly grouping with other Opalozoa but remained closely associated with eukaryotes although not with the metazoan enzymes. Although this fails to strengthen the case for a monophyletic origin of glycogen metabolism in Bigyra, it nevertheless does not reject it and further precludes possible LGTs from the animal hosts or from gut bacteria to the intestinal protist Blastocystis. Taken together, these results remain in agreement with a single ancient origin of alpha-glucan metabolism in Stramenopila.

Beta-Glucan Metabolism Displays a Vertical Inheritance in Gyrista

We performed phylogenetic analysis for the two enzymes that have been proven or are very highly suspected to be active in chrysolaminarin metabolism. We reveal that both the GT48 (**Figure 4**) and the GH16 subfamily 2 (**Figure 5** and **Supplementary Figure S3**) display a close relationship between Gyrista, Haptophyta, and Cercozoa. Indeed, in the GH16 subfamily 2 phylogeny, Haptophyta displays sisterhood and is even embedded inside the Stramenopila group as they are grouping with BS = 94 and pp = 0.94. In addition, the GT48 phylogeny possessed a somehow similar topology, with slight differences, the Cercozoa being the sister group (BS = 89, pp = 0.85) while the Haptophyta grouped with them with a BS of 78. Finally, we can observe in both

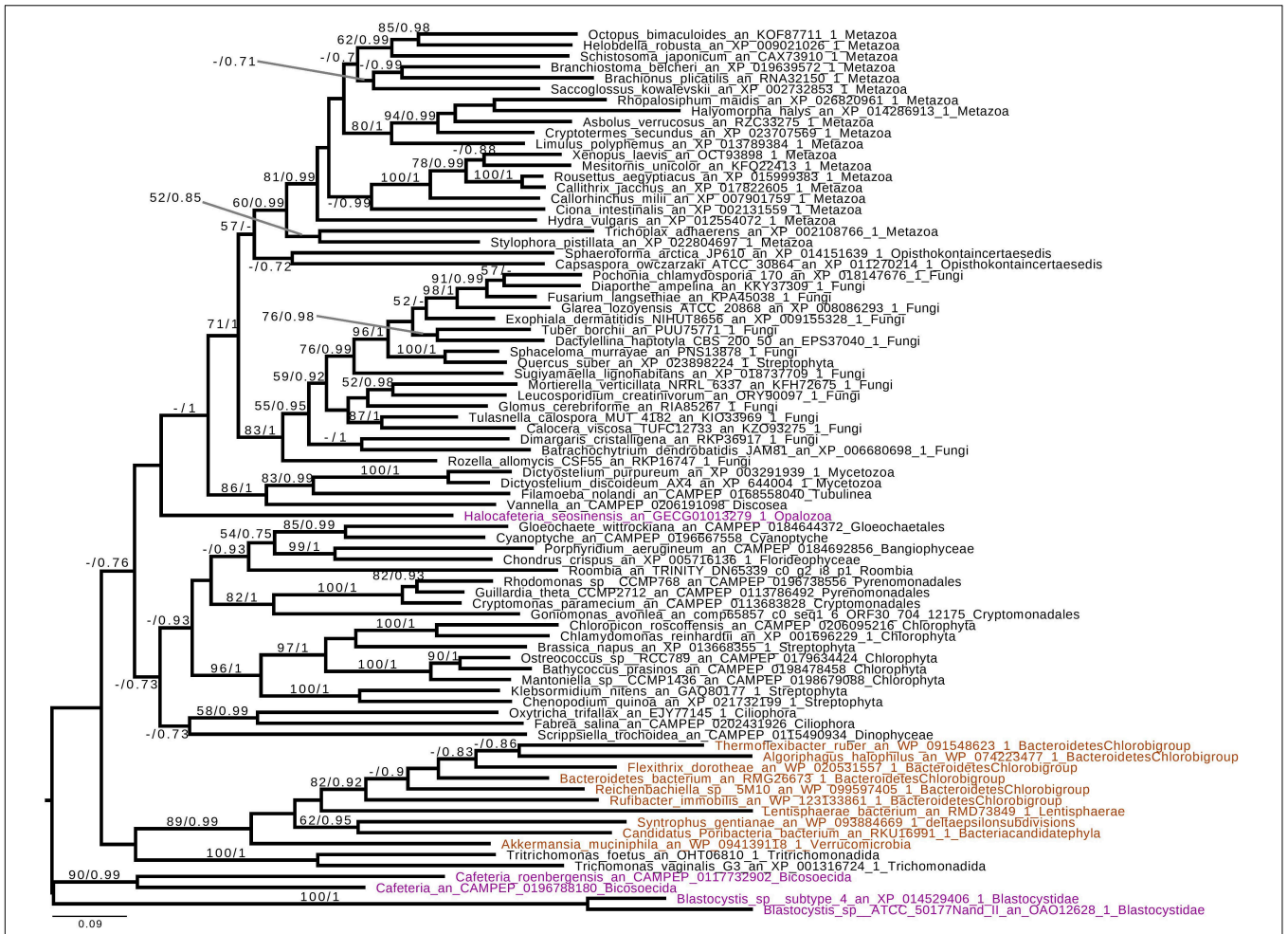


FIGURE 2 | Phylogenetic analysis of branching enzyme, Glycoside Hydrolase family 13 subfamily 8 (GH13_8). The tree displayed is midpoint rooted and represents the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values drawn from 100 bootstraps repetition with IQTREE (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50% are shown, while only posterior probabilities >0.6 are shown. The scale bar shows the inferred number of amino acid substitutions per site. Sequences are highlighted in purple for Stramenopila, brown for Bacteria, while everything else is in black. Sequence names are composed of the organism name, the accession number, and their clades. We can see that Blastocystis and Cafeteria group together, however, with a low bootstrap and posterior probabilities. In addition, Halocafeteria was not found inside a monophyletic group, but was found among other eukaryotes.

phylogenies two groups consisting of Fungi and Viridiplantae (Chlorophyta or Streptophyta, respectively, in GH16_2 and GT48 trees). In both cases, the enzymes are demonstrated to be used in cell wall metabolism rather than storage as beta-glucan storage polysaccharides (Bowen and Wheals, 2004; Nishikawa et al., 2005).

Nevertheless, they should have the same type of enzyme catalytic activity as in Stramenopila. It must be stressed that some Stramenopila also synthesize beta-glucans in their cell wall and that the presence of different Stramenopila subgroups in the phylogeny may be understood in this light. The GH81 phylogenetic tree (**Supplementary Figure S4**), and particularly part II, also displays the relation between Stramenopila, Cercozoa, and Haptophyta, but with a lower robustness (BS = 51 and pp = 0.88). However, the GH161 putative beta-glucan phosphorylase phylogenetic tree (**Supplementary Figure S5**) displays a totally different topology with Stramenopila and Dinoflagellate having exchanged their gene probably multiple times, while the sequence seems to originate from Bacteria. Finally, the GH5_33 (**Supplementary Figure S6**) enzyme displays a topology analogous to those of the GH16_2 and GT48 enzymes (with organisms from Haptophyta and Cercozoa), albeit with some additional complexities including the additional presence of Ochromonas, Ochromonadaceae, and ciliates on the top of the tree and the absence of plant sequences.

DISCUSSION

Putative New Enzymes of Beta-Glucan Metabolism

Beta-glucan storage polysaccharide metabolism remains an understudied pathway. Indeed, in Stramenopila, only two enzymes are known to be clearly involved in Chrysolaminarin

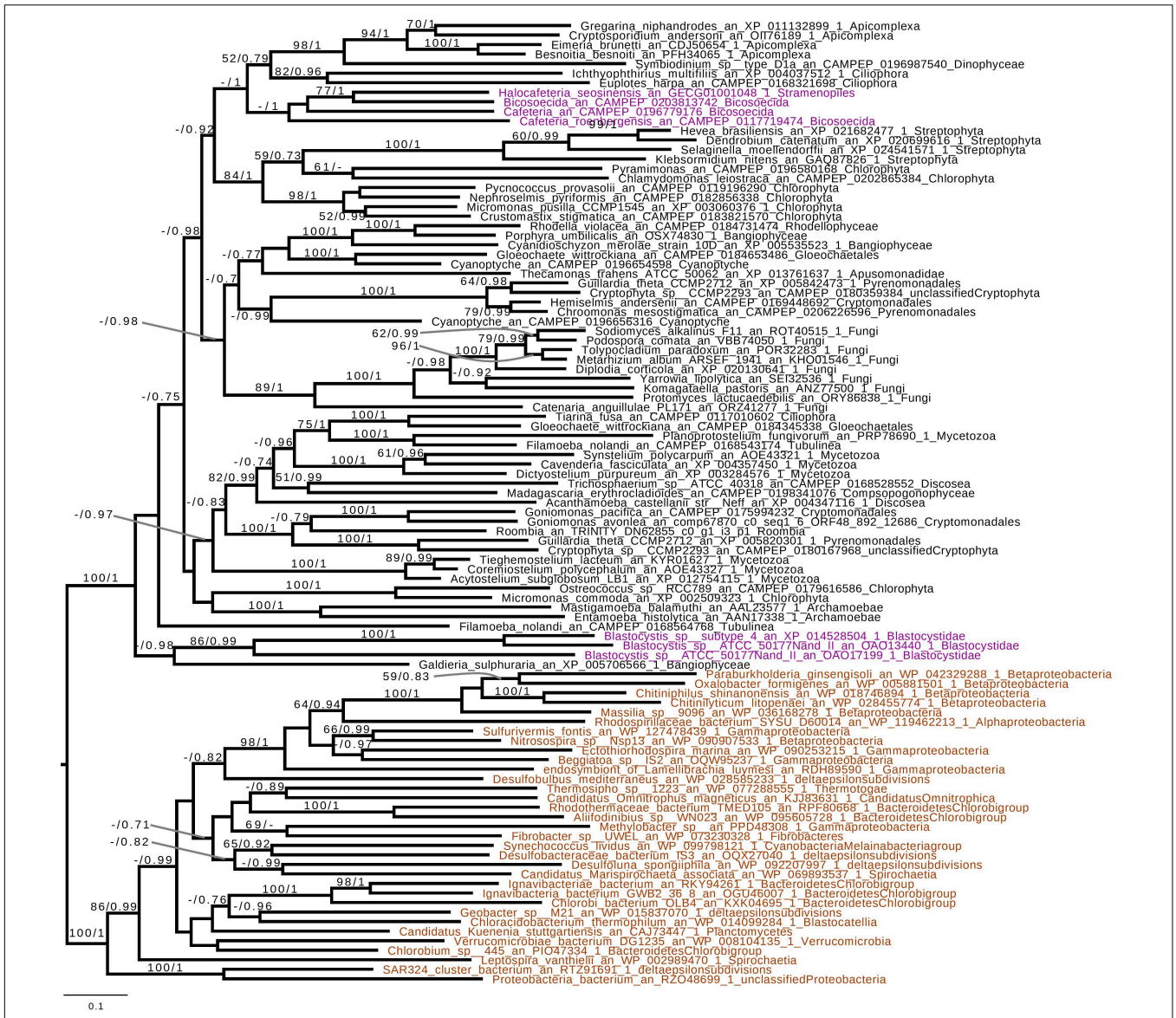


FIGURE 3 | Phylogenetic analysis of glycogen phosphorylase (GT35 family). The tree displayed is midpoint rooted and represents the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values drawn from 100 bootstraps repetition with IQTREE (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50% are shown, while only posterior probabilities >0.6 are shown. The scale bar shows the inferred number of amino acid substitutions per site. Sequences are highlighted in purple for Stramenopila, brown for Bacteria, while everything else is in black. Sequence names are composed of the organism name, the accession number, and their clades. We can see that Bicosoecida, Halocafeteria, and Cafeteria group together, with a high posterior probability (pp = 1). However, Blastocystis was not found inside a monophyletic group with Stramenopila but was found among other eukaryotes. The Blastocystis position is probably due to the fast-evolving sequence in this organism as it has been observed in several studies (Eme et al., 2017; Moreira and López-García, 2017).

biosynthesis, GT48 and GH16_2, and among these two, only GT48 was functionally fully demonstrated to be involved. By looking at the sequence distributions in Stramenopila using the CAZy database as basic knowledge to determine candidates involved in beta-glucan metabolism and the freely available HMM profile from the automated Carbohydrate-active enzyme annotation (dbCAN) to further assign the sequences detected to their corresponding enzyme subfamilies, we were able to propose the possible involvement of both the GH81 and GH5_33 enzymes. Interestingly, the GH81 phylogeny when concentrated around the Stramenopila group II (**Supplementary Figure S4**) displays a similar phylogenetic pattern to those seen with both GT48 and GH16_2, with the presence of Haptophyta and Cercozoa. Moreover, the GH5_33 phylogeny is also somewhat comparable to the topology of GT48 and GH16_2, thus for each enzyme (GH5_33 and GH81), the two correlations: on the one hand, co-occurrence of the enzymes (**Table 1**), and, on the other hand, the topology of their phylogeny suggests that they have evolved simultaneously and could be involved in chrysolaminarin catabolism.

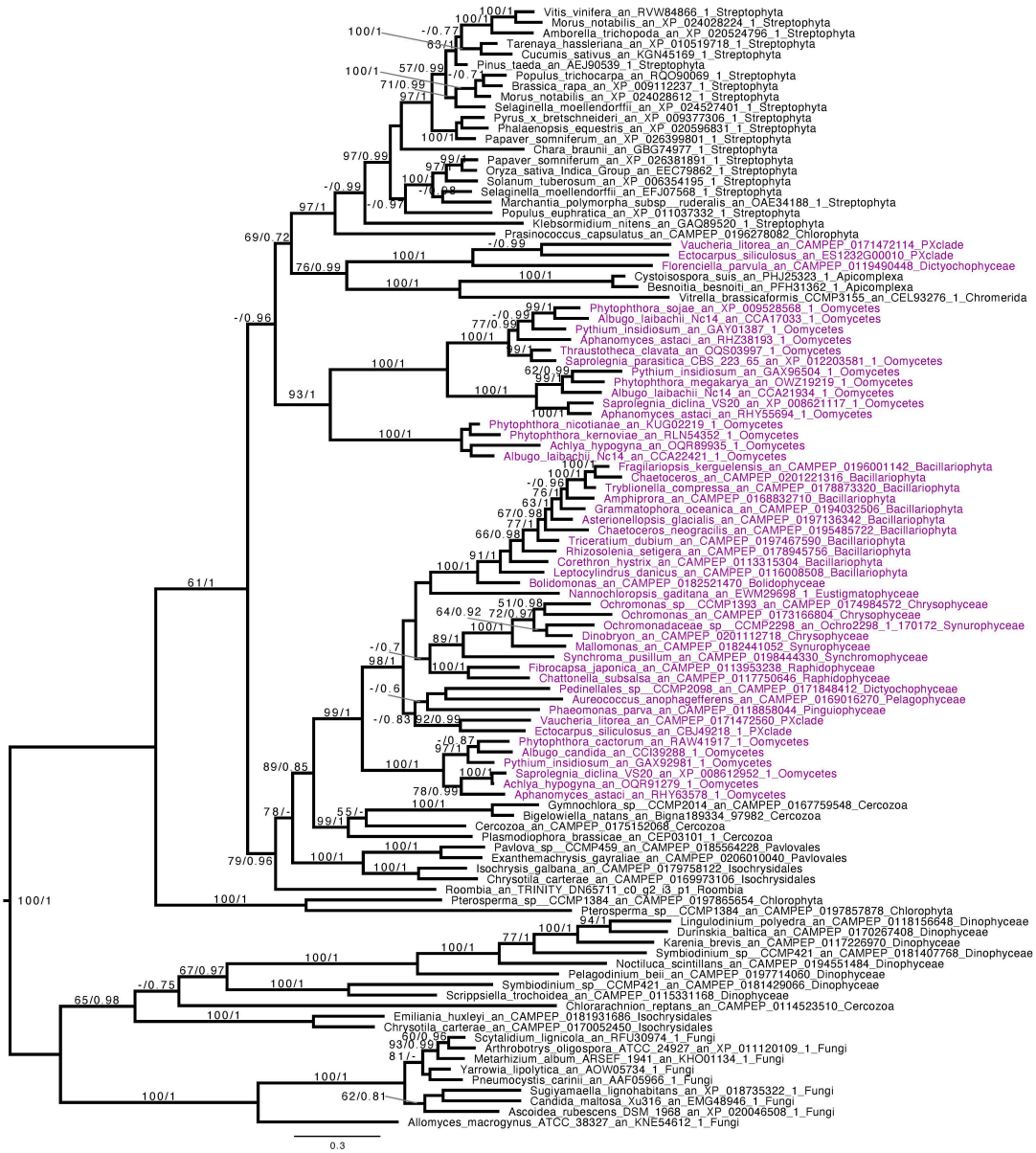


FIGURE 4 | Phylogenetic analysis of GT48. The tree displayed is midpoint rooted and represents the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values drawn from 100 bootstraps repetition with IQTREE (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values >50% are shown, while only posterior probabilities >0.6 are shown. The scale bar shows the inferred number of amino acid substitutions per site. Sequences are highlighted in purple for Stramenopila, while everything else is in black. Sequence names are composed of the organism name, the accession number, and their clades. We can observe a strongly supported group with all Gyrista (BS = 99, pp = 1), with inside the expected topology based on Stramenopila phylogeny. In addition, among those sequences we find the characterized enzyme from *P. tricornutum*. Then, this group of sequence is likely to be the one involved in laminarin biosynthesis. Moreover, we can find close to this Gyrista group several sequences from both Haptophyta and Cercozoa (BS = 79, pp = 0.96), probably also involved in storage polysaccharide metabolism.

In addition, phosphorolysis of storage polysaccharides was recently inferred to play a role in beta-glucan catabolism by generating glucose-1-phosphate, thereby retaining some of the free energy required for synthesis. We report here the distribution in Stramenopila of the new GH161 family (Kuhaudomlarp et al., 2019; **Table 1**), which could be involved in Chrysolaminarin catabolism. Surprisingly, this enzyme seems to be only present in Ochrophyta. In diatoms, it was proposed by Kroth et al. (2008) that absence of phosphorolysis was compensated by the presence of glucokinase; however, because of the presence of GH161, this could very well not be entirely true. However, the absence of GH161 in oomycetes suggests that the presence of such an enzyme could be dispensable and glucokinase, for instance, could be sufficient to generate

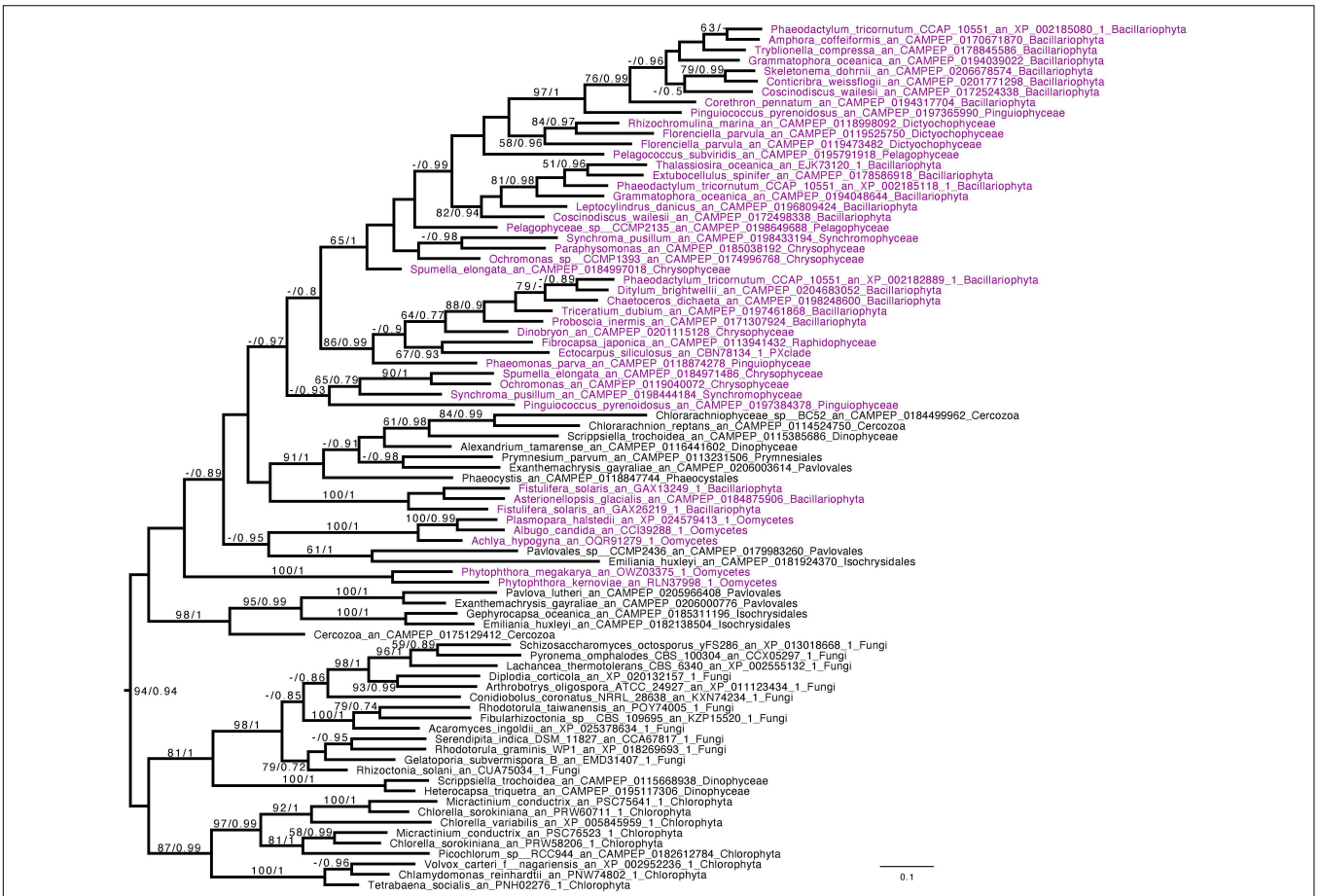


FIGURE 5 | Phylogenetic analysis of GH16_2. The tree displayed is manually rooted according to the topology obtained in **Supplementary Figure S3**. It represents the consensus tree obtained with Phylobayes 4.1 with ML bootstrap values drawn from 100 bootstraps repetition with IQTREE (left) and Bayesian posterior probabilities (right) mapped onto the nodes. Bootstrap values (BS) >50 are shown, while only posterior probabilities (pp) >0.6 are shown. The scale bar shows the inferred number of amino acid substitutions per site. Sequences are highlighted in purple for Stramenopila, while everything else is in black. Sequences names are composed of the organism name, the accession number, and their clades. GH16_2 from the Stramenopila group together with sequences from Haptista and Cercozoa with a BS = 94 and a pp = 0.94, mirroring the topology from the GT48 phylogenetic tree.

the required sugar phosphate. This result also suggests that GH161 may have been acquired in an endosymbiotic context and is probably favored by the presence of glycolysis in two different compartments (i.e., cytosol and plastid). Moreover, GH161 sequences from the Ochrophyta group with bacterial sequences, which suggest a lateral gene transfer from Bacteria to an Ochrophyta ancestor. Finally, even if we did not find a sequence in *E. siliculosus* (**Supplementary Figure S5**), the presence of GH161 enzymes in the brown alga sister group Raphidophyceae (Fibrocapsa) suggests that the gene was either simply not found or lost in *Ectocarpus*, further suggesting the dispensable nature of phosphorolysis in beta-glucan metabolism.

Possible Relative Merits of Alpha- vs. Beta-Glucan Water-Soluble Polysaccharides

While soluble alpha-glucan storage polysaccharides are always synthesized in the cytosol and degraded in both the cytosol and lysosome, the soluble beta-glucan storage polysaccharides such as Chrysolaminarin are synthesized, kept, and degraded in dedicated vacuoles. A rationale behind this observation may be found in the structure of those polysaccharides; indeed, laminarin displays a molecular weight varying from 3 to 10 kDa (Kadam et al., 2015) while that of glycogen is usually between 10^3 and 10^4 kDa (Sullivan et al., 2012), involving very important differences (up to three orders of magnitude) in their respective degree of glucose polymerization, which directly impacts the osmotic pressure of the cellular compartment where they are synthesized. Thereby, the high degree of polymerization of glycogen makes it way more osmotically tolerable for the fragile cytosol of eukaryotes while Chrysolaminarin-type polysaccharide/oligosaccharides can represent, if in high amounts, a problem that is, at least, partly resolved by its vacuolar localization. The penalties and benefits ensuing from storage compounds of, respectively, high and low degree of polymerization will thus vary greatly with the presence or absence of a rigid vegetative cell wall and with

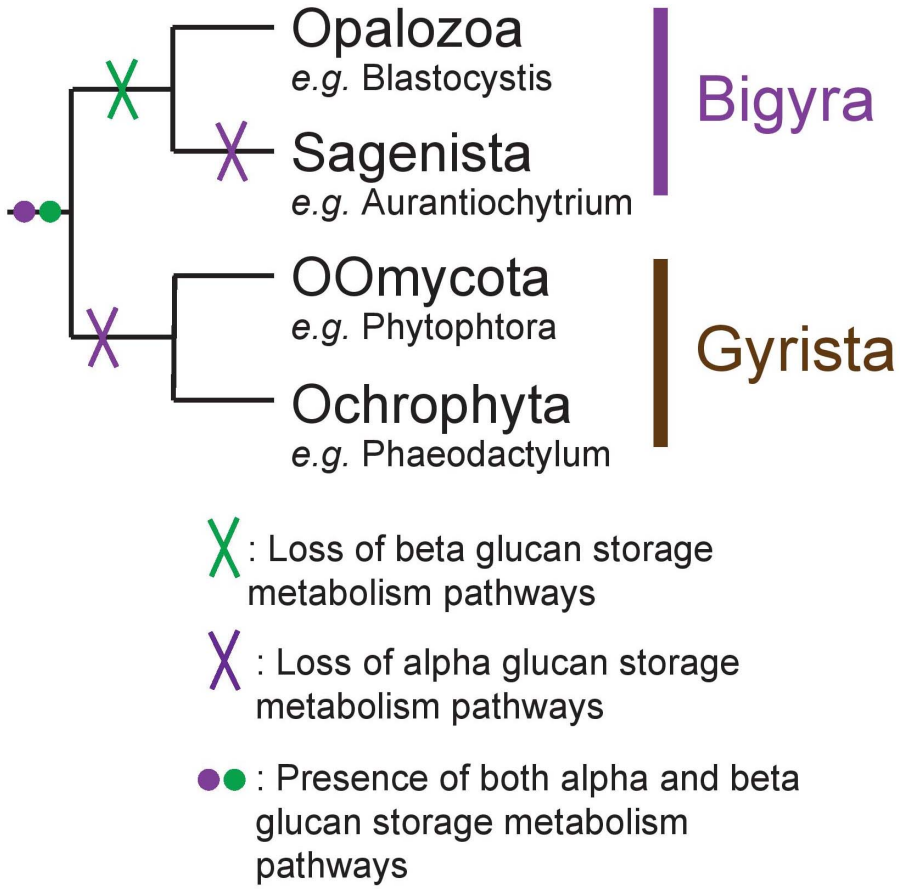


FIGURE 6 | Schematic representation of the Stramenopila evolution tree with the hypothesis defended here with the presence of both alpha- and beta-glucan storage polysaccharide metabolisms in the ancestor, represented by the purple and green dot, respectively. Their respective loss is indicated by a cross on the branch: in Bigyra, the loss of beta-glucan metabolism is indicated by a green cross, while alpha-glucan metabolism has been lost twice in Gyrista lineages as well as in Sagenista (purple cross).

the subcellular localization of such compounds. Despite these considerations, it remains, however, very hard to understand what has apparently favored the selection and replacement of alpha-glucan metabolism by that of beta-glucans in a majority of rhizarians and most stramenopiles over the maintenance of glycogen metabolism in all alveolates and opalozoans.

Origin of Metabolism in Stramenopila

If we remove storage lipid metabolism from our reasoning and restrict our comparisons to alpha- and beta-glucan metabolism as carbon storage material, we observe that those pathways seem mutually exclusive in Stramenopila as they are in all eukaryotes. While beta-glucan is found in Gyrista, alpha-glucan is found in Opalozoan which begs the question of the ancestrality of both types of metabolism. To shed light on such issues, we focused our attention to our phylogenetic analysis. This shows that in the alpha-glucan metabolism iDBE (GH13_25-GH133) from Stramenopila (**Supplementary Figure S1**) seems to form a cluster with other Alveolata, the closest related group from Stramenopila (with low statistical value pp = 0.7, BS < 50). The same observation can be made for the GT35 phylogeny of Halocafeteria and Cafeteria (**Figure 3**) (pp = 1), also suggesting a vertical inheritance. We should note that for GT35, Blastocystis is clustering more distantly, which could be due to the fast-evolving sequence in this organism (Eme et al., 2017; Moreira and López-García, 2017). In addition, both GH13_8 (**Figure 2**) and GT5 (**Supplementary Figure S2**) are found close to other eukaryotic sequences. Altogether, these phylogenies suggest that the Stramenopila ancestors initially had this ancient eukaryotic metabolism.

Beta-glucan metabolism shows a common pattern between GT48 (**Figure 4**), GH16_2 (**Figure 5**), and GH81 part II (**Supplementary Figure S4**), with a clustering with Haptophyta and Cercozoa. In addition, the GH5_33 phylogenies display a related topology (**Supplementary Figure S6**), while the group on the top is composed of both Ochromonas and Ochromonadaceae with ciliates that are associated with a long branch that could indicate another function in this particular cluster. On the contrary, GH161 (**Supplementary Figure S5**) seems to have been acquired specifically in the Ochrophyta ancestor by lateral gene transfer from Bacteria.

These results on beta-glucan storage metabolism in Stramenopila might also suggest an ancient relation between Stramenopila and beta-glucan polysaccharide. In particular, since

Rhizaria and Haptista are closely related groups to Stramenopila in eukaryote phylogeny (Burki et al., 2016). A single evolution of storage beta-glucan metabolism at the base of the SAR group followed by segregation of alpha- and beta-glucan metabolisms leading to maintenance of glycogen and loss of beta-glucans in alveolate and opalozoa can at least be proposed.

Alternatively, the fact that, among Stramenopiles, only Gyrista can store beta-glucan might indicate that this metabolism could have been gained through the endosymbiosis process of Haptophyta as it has been proposed in the serial hypothesis (Baurain et al., 2010). In this case, however, we would have to infer that Oomycetes originated after the secondary endosymbiosis event that would have generated the Gyrista.

Altogether, these results point toward an acquisition of storage beta-glucan polysaccharide metabolism through gene inheritance rather than from a specific metabolic sub-functionalization that occurred specifically in Stramenopila. This result suggests that re-targeting of cell wall metabolism, as proposed in Kroth et al. (2008), to storage polysaccharide metabolism did not occur in Stramenopila, but could have nevertheless occurred in other organisms.

Ancient Evolution of Beta and Alpha Storage Polysaccharide Metabolisms

Understanding storage polysaccharide metabolism distribution in Stramenopiles requires a better understanding of their ancestral status. To address this question, we investigated the metabolic distribution of alpha- and beta-glucans in eukaryotes. Interestingly, alpha-glucan storage polysaccharides seem to be found in all eukaryotes: unikonts (Amoebozoan, Obazoa) and bikonts (SAR, Archaeplastida, Cryptista, Haptista, Excavata) alike (Ball et al., 2011). In particular, this work shows a very ancient relation between Stramenopila and alpha-glucan storage metabolism, while soluble beta-glucan storage polysaccharides seem to be only found in some bikonts and never found in unikonts.

In addition, although cell wall beta-1,3 glucan polysaccharides are widespread in bikonts and unikonts, the beta-1,3-glucan storage polysaccharides are limited to bikonts, suggesting a scenario whereby beta-glucan storage metabolism may have evolved from preexisting cell wall polysaccharide metabolism at the base of the SAR and haptophytes while alpha-glucan storage metabolism evolved during eukaryogenesis and thus pre-existed in both the ancestor of the SAR and that of the haptophytes. Hence, ancestors of such organisms could have possessed both alpha- and beta-glucan storage metabolisms. Indeed, this exceptional situation seems to exist in both Goniomonas (Cenci et al., 2018b) and Emiliana (data not published).

While we acknowledge the fact that scarcity of genomes in Opalozoa could induce biases in our interpretation, we propose nevertheless that, in Stramenopila, there was an ancient presence of both alpha- and beta-glucan storage polysaccharide metabolism (**Figure 6**). This was followed by loss of beta-glucan storage metabolism in some lineages and of loss alpha-glucan metabolism in others while a few storage lipid accumulating organisms lost both.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s. Publicly available datasets were analyzed in this study. Phylogenetic datasets analyzed are available at https://doi.org/10.5061/dryad.dv41ns1x3.

AUTHOR CONTRIBUTIONS

UC and SB designed the project. MCh and UC performed the annotation. MCh, ML, LF, and UC performed the phylogenetic analysis. MCh, ML, LF, MCo, CC, SB, and UC analyzed the results. UC and SB wrote the manuscript. All authors read and accepted the final manuscript.

FUNDING

This work was supported by the CNRS, the Université de Lille, the ANR grant "Expendo" (ANR-14-CE11-0024). In addition, ML was financed by the ANR grant "MathTest" (ANR-18-CE13-0027), and MCh by the European Regional Development Fund "Alibiotech."

ACKNOWLEDGMENTS

We thank Connie Lovejoy (Université Laval) for the advance provision of genome sequence data from *Ochromonadaceae* sp. CCMP2298. This sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community (Project ID: 1014525).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.629045/full#supplementary-material

Supplementary Figure 1 | Phylogenetic analysis of GH13_25 (iDBE).

Supplementary Figure 2 | Phylogenetic analysis of GT5.

Supplementary Figure 3 | Phylogenetic analysis of GH16_2 rooted with other GH16 sequences.

Supplementary Figure 4 | Phylogenetic analysis of GH81.

Supplementary Figure 5 | Phylogenetic analysis of GH161.

Supplementary Figure 6 | Phylogenetic analysis of GH5_33.

Supplementary Table 1 | Table of CAZy families and subfamilies found in 17 Stramenopila genomes, with the accession number for each of the annotated sequences.

Supplementary Table 2 | Table of targeting prediction and transmembrane domains for GT48, GH16_2, GH81, GH5_33, and GH161.

REFERENCES

Almagro Armenteros, J. J., Salvatore, M., Emanuelsson, O., Winther, O., von Heijne, G., Elofsson, A., et al. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2, e201900429. doi: 10.26508/lsa.201900429

Ball, S., Colleoni, C., Cenci, U., Raj, J. N., and Tirtiaux, C. (2011). The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J. Exp. Bot.* 62, 1775–1801. doi: 10.1093/jxb/erq411

Baurain, D., Brinkmann, H., Petersen, J., Rodriguez-Ezpeleta, N., Stechmann, A., Demoulin, V., et al. (2010). Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* 27, 1698–1709. doi: 10.1093/molbev/msq059

Bowen, S., and Wheals, A. (2004). Incorporation of Sed1p into the cell wall of involves. *FEMS Yeast Res.* 4, 731–735. doi: 10.1016/j.femsyr.2004.02.006

Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B. Q., Radaykina, L. V., et al. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B Biol. Sci.* 283, 20152802. doi: 10.1098/rspb.2015.2802

Cavalier-Smith, T. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* 46, 347–366. doi: 10.1111/j.1550-7408.1999.tb04614.x

Cenci, U., Moog, D., Curtis, B. A., Tanifuji, G., Eme, L., Lukeš, J., et al. (2016). Heme pathway evolution in kinetoplastid protists. *BMC Evol. Biol.* 16:109. doi: 10.1186/s12862-016-0664-6

Cenci, U., Nitschke, F., Steup, M., Minassian, B. A., Colleoni, C., and Ball, S. G. (2014). Transition from glycogen to starch metabolism in Archaeplastida. *Trends Plant Sci.* 19, 18–28. doi: 10.1016/j.tplants.2013.08.004

Cenci, U., Qiu, H., Pillonel, T., Cardol, P., Remacle, C., Colleoni, C., et al. (2018a). Host-pathogen biotic interactions shaped vitamin K metabolism in Archaeplastida. *Sci. Rep.* 8:15243. doi: 10.1038/s41598-018-33663-w

Cenci, U., Sibbald, S. J., Curtis, B. A., Kamikawa, R., Eme, L., Moog, D., et al. (2018b). Nuclear genome sequence of the plastid-lacking cryptomonad Goniomonas avonlea provides insights into the evolution of secondary plastids. *BMC Biol.* 16:137. doi: 10.1186/s12915-018-0593-5

Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R., et al. (2014). Chromosome scale genome assembly and transcriptome profiling of nannochloropsis gaditana in nitrogen depletion. *Mol. Plant* 7, 323–335. doi: 10.1093/mp/sst120

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* 10:210. doi: 10.1186/1471-2148-10-210

Derelle, R., López-García, P., Timpano, H., and Moreira, D. (2016). A phylogenomic framework to study the diversity and evolution of stramenopiles (= Heterokonts). *Mol. Biol. Evol.* 33, 2890–2898. doi: 10.1093/molbev/msw168

Dinadayala, P., Sambou, T., Daffe, M., and Lemassu, A. (2008). Comparative structural analyses of the -glucan and glycogen from *Mycobacterium bovis*. *Glycobiology* 18, 502–508. doi: 10.1093/glycob/cwn031

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M., and Roger, A. J. (2017). Lateral gene transfer in the adaptation of the anaerobic parasite blastocystis to the gut. *Curr. Biol.* 27, 807–820. doi: 10.1016/j.cub.2017.02.003

Field, C. B. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Gentekaki, E., Curtis, B. A., Stairs, C. W., Klimeš, V., Eliáš, M., Salas-Leiva, D. E., et al. (2017). Extreme genome diversity in the hyper-prevalent parasitic eukaryote Blastocystis. *PLoS Biol.* 15:e2003769. doi: 10.1371/journal.pbio.2003769

Graiff, A., Ruth, W., Kragl, U., and Karsten, U. (2016). Chemical characterization and quantification of the brown algal storage compound laminarin — a new methodological approach. *J. Appl. Phycol.* 28, 533–543. doi: 10.1007/s10811-015-0563-z

Gruber, A., and Kroth, P. G. (2017). Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160402. doi: 10.1098/rstb.2016.0402

Harding, T., Brown, M. W., Simpson, A. G. B., and Roger, A. J. (2016). Osmoadaptative strategy and its molecular signature in obligately halophilic heterotrophic protists. *Genome Biol. Evol.* 8, 2241–2258. doi: 10.1093/gbe/evw152

Huang, W., Haferkamp, I., Lepetit, B., Molchanova, M., Hou, S., Jeblick, W., et al. (2018). Reduced vacuolar β-1,3-glucan synthesis affects carbohydrate metabolism as well as plastid homeostasis and structure in *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 4791–4796. doi: 10.1073/pnas.1719274115

Huang, W., Río Bártulos, C., and Kroth, P. G. (2016). Diatom vacuolar 1,6-β-Transglycosylases can functionally complement the respective yeast mutants. *J. Eukaryot. Microbiol.* 63, 536–546. doi: 10.1111/jeu.12298

Kadam, S. U., Tiwari, B. K., and O'Donnell, C. P. (2015). Extraction, structure and biofunctional activities of laminarin from brown algae. *Int. J. Food Sci. Technol.* 50, 24–31. doi: 10.1111/ijfs.12692

Kamoun, S. (2003). Molecular genetics of pathogenic oomycetes. *Eukaryot. Cell* 2, 191–199. doi: 10.1128/EC.2.2.191-199.2003

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keeling, P., Burki, F., Wilcox, J., and Allam, B. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Kitaoka, M., Matsuoka, Y., Mori, K., Nishimoto, M., and Hayashi, K. (2012). Characterization of a bacterial laminaribiose phosphorylase. *Biosci. Biotechnol. Biochem.* 76, 343–348. doi: 10.1271/bbb.110772

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., et al. (2008). A model for carbohydrate metabolism in the diatom phaeodactylum tricornutum deduced from comparative whole genome analysis. *PLoS One* 3:e1426. doi: 10.1371/journal.pone.0001426

Kuhaudomlarp, S., Patron, N. J., Henrissat, B., Rejzek, M., Saalbach, G., and Field, R. A. (2018). Identification of Euglena gracilis β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149. *J. Biol. Chem.* 293, 2865–2876. doi: 10.1074/jbc.RA117.000936

Kuhaudomlarp, S., Pergolizzi, G., Patron, N. J., Henrissat, B., and Field, R. A. (2019). Unraveling the subtleties of β-(1→3)-glucan phosphorylase specificity in the GH94, GH149 and GH161 glycoside hydrolase families. *J. Biol. Chem.* 294, 6483–6493. doi: 10.1074/jbc.RA119.007712

Lee Chang, K. J., Dunstan, G. A., Abell, G. C. J., Clementson, L. A., Blackburn, S. I., Nichols, P. D., et al. (2012). Biodiscovery of new Australian thraustochytrids for production of biodiesel and long-chain omega-3 oils. *Appl. Microbiol. Biotechnol.* 93, 2215–2231. doi: 10.1007/s00253-011-3856-4

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178

Maruyama, S., Eveleigh, R. J., and Archibald, J. M. (2013). Treetrimmer: a method for phylogenetic dataset size reduction. *BMC Res. Notes* 6:145. doi: 10.1186/1756-0500-6-145

Michel, G., Tonon, T., Scornet, D., Cock, J. M., and Kloareg, B. (2010). The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes. *New Phytol.* 188, 82–97. doi: 10.1111/j.1469-8137.2010.03374.x

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263

Moreira, D., and López-García, P. (2017). Protist evolution: stealing genes to gut it out. *Curr. Biol.* 27, R223–R225. doi: 10.1016/j.cub.2017.02.010

Mouille, G., Maddelein, M.-L., Libessart, N., Talaga, P., Decq, A., Delrue, B., et al. (1996). Preamylopectin processing: a mandatory step for starch biosynthesis in plants. *Plant Cell* 8, 1353–1366. doi: 10.2307/3870306

Nishikawa, S., Zinkl, G. M., Swanson, R. J., Maruyama, D., and Preuss, D. (2005). Callose (beta-1,3 glucan) is essential for *Arabidopsis* pollen wall patterning, but not tube growth. *BMC Plant Biol.* 5:22. doi: 10.1186/1471-2229-5-22

O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R. G., Hamilton, C. J., et al. (2015). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Mol. BioSyst.* 11, 2808–2820. doi: 10.1039/C5MB00319A

Petsalaki, E. I., Bagos, P. G., Litou, Z. L., and Hamodrakas, S. J. (2006). PredSL: A Tool for the N-terminal Sequence-based Prediction Protein Subcellular Localization. *Genom. Proteom. Bioinform.* 4, 48–55. doi: 10.1016/S1672-0229(06)60016-8

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Stiller, J. W., Schreiber, J., Yue, J., Guo, H., Ding, Q., and Huang, J. (2014). The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* 5:5764. doi: 10.1038/ncomms6764

Sullivan, M. A., O'Connor, M. J., Umana, F., Roura, E., Jack, K., Stapleton, D. I., et al. (2012). Molecular insights into glycogen α-particle formation. *Biomacromolecules* 13, 3805–3813. doi: 10.1021/bm3012727

Teste, M. A., Enjalbert, B., Parrou, J. L., and FranÃ§ois, J. M. (2000). The *Saccharomyces cerevisiae YPR184w* gene encodes the glycogen debranching enzyme. *FEMS Microbiol. Lett.* 193, 105–110. doi: 10.1111/j.1574-6968.2000.tb09410.x

Viborg, A. H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., et al. (2019). A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* 294, 15973–15986. doi: 10.1074/jbc.RA119.010619

Vlierberghe, M. V., Léonard, R. R., Cornet, L., Bouzin, Y., and Baurain, D. (2021). Curated sequence database "Life-OF-Mick" first used in Cornet, Magain et al. Exploring syntenic conservation across genomes for phylogenetic studies of organisms subjected to horizontal gene transfers: a case study with Cyanobacteria. doi: 10.6084/M9.FIGSHARE.13550267.V2

Wilson, W. A., Roach, P. J., Montero, M., Baroja-Fernández, E., Muñoz, F. J., Eydallin, G., et al. (2010). Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol. Rev.* 34, 952–985. doi: 10.1111/j.1574-6976.2010.00220.x

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

Yoshimi, A., Miyazawa, K., and Abe, K. (2017). Function and biosynthesis of cell wall α-1,3-Glucan in fungi. *J. Fungi* 3:63. doi: 10.3390/jof3040063

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chabi, Leleu, Fermont, Colpaert, Colleoni, Ball and Cenci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 12 April 2021
doi: 10.3389/fpls.2021.638181

Comparative Structural and Functional Analyses of the Fusiform, Oval, and Triradiate Morphotypes of *Phaeodactylum tricornutum* Pt3 Strain

Ludovic Galas[1†], Carole Burel[2†], Damien Schapman[1], Marc Ropitaux[2], Sophie Bernard[1,2], Magalie Bénard[1] and Muriel Bardor[2,3*]*

[1] Normandie University, UNIROUEN, INSERM, PRIMACEN, Rouen, France, [2] Normandie University, UNIROUEN, Laboratoire Glycobiologie et Matrice Extracellulaire Végétale (Glyco-MEV) EA4358, Rouen, France, [3] Institut Universitaire de France, Paris, France

The diatom *Phaeodactylum tricornutum* is a marine unicellular microalga that exists under three main morphotypes: oval, fusiform, and triradiate. Previous works have demonstrated that the oval morphotype of *P. tricornutum* Pt3 strain presents specific metabolic features. Here, we compared the cellular organization of the main morphotypes of the diatom *P. tricornutum* Pt3 strain through transmission electron and advanced light microscopies. The three morphotypes share similarities including spectral characteristics of the plastid, the location of the nucleus, the organization of mitochondria around the plastid as well as the existence of both a F-actin cortex, and an intracellular network of F-actin. In contrast, compared to fusiform and triradiate cells, oval cells spontaneously release proteins more rapidly. In addition, comparison of whole transcriptomes of oval versus fusiform or triradiate cells revealed numerous differential expression of positive and negative regulators belonging to the complex dynamic secretory machinery. This study highlights the specificities occurring within the oval morphotype underlying that the oval cells secrete proteins more rapidly.

Keywords: microalgae, diatom, morphotype, organelles, cytoskeleton, secretion, *Phaeodactylum tricornutum*, biofactory

OPEN ACCESS

Edited by:
Benoit Schoefs, Le Mans Université, France

Reviewed by:
Inna Khozin-Goldberg, Ben-Gurion University of the Negev, Israel
Tore Brembu, Norwegian University of Science and Technology, Norway
Fabrice Franck, University of Liège, Belgium

***Correspondence:**
Muriel Bardor
muriel.bardor@univ-rouen.fr
Ludovic Galas
ludovic.galas@univ-rouen.fr

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *05 December 2020*
Accepted: *16 March 2021*
Published: *12 April 2021*

Citation:
Galas L, Burel C, Schapman D, Ropitaux M, Bernard S, Bénard M and Bardor M (2021) Comparative Structural and Functional Analyses of the Fusiform, Oval, and Triradiate Morphotypes of Phaeodactylum tricornutum Pt3 Strain. Front. Plant Sci. 12:638181. doi: 10.3389/fpls.2021.638181

INTRODUCTION

During the last decades, knowledge regarding cell biology of eukaryotic model organisms like plants, yeast, animal cells have increased tremendously (Bezanilla, 2013; Martin, 2014; Mathur et al., 2017). In contrast, comprehension of cellular processes from the marine diatom *Phaeodactylum tricornutum* is still limited. *P. tricornutum* is an unicellular Stramenopile believed to have arisen *via* a serial endosymbiotic event in which a red microalga were engulfed by a heterotroph (Moustafa et al., 2009; Bowler et al., 2010; Prihoda et al., 2012), thus generating specific genomic features and metabolic pathways (Bowler et al., 2008; Keeling and Palmer, 2008). Indeed, a recent investigation of *P. tricornutum* genome revealed that a total of 3,170 genes (26%) are unique and specific to this organism (Rastogi et al., 2018). *P. tricornutum* also harbors a combination of genes and metabolic

pathways that belongs either to the plant or animal kingdoms (C4 photosynthetic pathway and urea cycle, for example) (Butler et al., 2020). It is a photoautotrophic organism for which molecular tools as well as transformation methods have been developed (Apt et al., 1996; Zaslavskaia et al., 2000; Niu et al., 2012; Miyahara et al., 2013; Zhang and Hu, 2014; Karas et al., 2015). Indeed, genetic engineering such as gene silencing (De Riso et al., 2009), TALEN (Daboussi et al., 2014; Serif et al., 2017), and CRISPR/cas9 (Nymark et al., 2016; Mann et al., 2017; Serif et al., 2018; Slattery et al., 2018; Stukenberg et al., 2018) has been proven to be efficient in *P. tricornutum*. These tools should in the near future help in deciphering cellular processes and optimizing the commercial exploitation of *P. tricornutum*, which naturally synthesizes numerous compounds of interest like pigments and omega-3 (Cadoret et al., 2012; Sasso et al., 2012; Hamilton et al., 2015; Kuczynska et al., 2015; Butler et al., 2020). In addition, *P. tricornutum* has been recently used for biotechnological applications such as the production of biopharmaceuticals including monoclonal antibodies (mAbs) (León-Bañares et al., 2004; Mathieu-Rivet et al., 2014; Hempel and Maier, 2016; Dumontier et al., 2018; Rosales-Mendoza et al., 2020). For example, engineered *P. tricornutum* is able to produce recombinant human anti-Marburg virus mAbs (Hempel et al., 2017) and functionally glycosylated human anti-hepatitis B mAbs (Hempel et al., 2011; Hempel and Maier, 2012; Vanier et al., 2015, 2018). Even successful, industrial exploitation and commercialization are still rather limited due to the amount of mAbs produced. Thus, increasing the production yield is a prerequisite before any industrialization of algae-made mAbs. Such improvement requires a better comprehension of the cellular and metabolism processes as well as the secretion mechanism. In the future, metabolic engineering strategies could be envisioned and implemented as exemplified for the production of high-value plant triterpenoid production (D'Adamo et al., 2019) and increase of lipid accumulation (Zou et al., 2018).

Phaeodactylum tricornutum is atypical as it occurs naturally in at least three distinct morphotypes: oval, fusiform, and triradiate (Borowitzka and Volcani, 1978). A fourth cruciform morphotype has sometimes being described (Wilson, 1946; Lewin et al., 1958; He et al., 2014). However, the fusiform morphotype is the more frequent one. It can be morphologically transformed under specific culture conditions into the oval or triradiate one, morphotypes being able eventually to switch back to the fusiform morphotype (Borowitzka and Volcani, 1978). Such plasticity is likely to be due to the poorly silicified cell wall of the fusiform morphotype. In contrast, the oval morphotype contains organized silicified frustules (Vartanian et al., 2009). Indeed, *P. tricornutum* cell is encased by a rigid silica frustule comprised of two overlapping thecae (Epithecae and Hypothecae), each composed of a valve and accompanying the girdle band (GB) region. The raphe represents slits of the valves allowing the secretion of mucilage that is involved in cell motility and adhesion (Martin-Jézéquel and Tesson, 2013; Willis et al., 2013). Recently, a pairwise comparison of the transcriptomes of the three morphotypes from *P. tricornutum* Pt3 strain revealed that 1% of genes are differentially expressed between the fusiform and the triradiate mophotypes whereas more than 22 and 29% are differentially expressed when comparing the oval versus fusiform and the oval versus triradiate, respectively (Ovide et al., 2018). Among the differentially expressed genes in the oval morphotype, genes encoding proteins involved in stress responses like heat shock proteins and protein containing DER1-like domain are up-regulated (Ovide et al., 2018). Such results agreed with previous observations which conclude that the oval morphotype represent a resistance form to stresses and survive in unfavorable conditions such as hyposaline conditions, low temperature, low light (Gutenbrunner et al., 1994; De Martino et al., 2007, 2011; Bartual et al., 2008). In agreement, it has recently been reported that 68% of the differentially expressed genes compared to the other morphotypes were found to be up-regulated and involved in the biosynthesis of triglyceride, glucuronomannan and nucleotide pathways (Ovide et al., 2018). In addition, these RNA-Seq data suggest that several components of the secretory machinery are regulated in the oval morphotype suggesting specific protein release (Ovide et al., 2018). In this work, we compare the structural features, cellular organization and kinetics of protein release of the three main morphotypes of *P. tricornutum,* namely the fusiform, oval and triradiate.

MATERIALS AND METHODS

Culture and Growth Conditions of *Phaeodactylum tricornutum*

Fusiform, oval, or triradiate morphotype enriched cultures of *P. tricornutum* Pt3 strain (CCAP 1052/1B; CCMP 2558) were generated as previously described (Ovide et al., 2018). *P. tricornutum* cells were grown at 19°C in 1 L flask on a 16/8 h light/dark cycle with light intensity of 68 μmol m^{-2} s^{-1}. The nutritive medium was composed of 100% seawater (Instant Ocean) for the fusiform and triradiate morphotypes and of 10% seawater (Instant Ocean) for the oval morphotype. Sterilized by filtration through a 0.22 μm filter and autoclaved, seawater was then complemented with trace elements and 80 mg L^{-1} of sodium metasilicate (Na_2SiO_3) as previously reported (Baïet et al., 2011). The diatom cells were cultured under ambient air. CO_2 from the air was the only available source of carbon.

Ultrastructural Characterization of *P. tricornutum* Morphotypes Through Transmitted Electron Microscopy

High pressure freezing (HPF) was performed with the HPF-EM PACT I freezer from Leica Microsystems (Nanterre, France). Prior to freezing, cells were treated with 100 mM mannitol during 2 h at room temperature for cryopreservation. Pre-treated diatoms were then transferred into the cavity of a copper ring (diameter of 1.2 mm; depth of 100 μm). Using a horizontal loading station, the specimen carriers were tightened securely to the pod of specimen holder. After fixation on the loading device, specimen were frozen with a maximum cooling rate of 10,000°C s^{-1}, an incoming pressure of 7.5 bars and a working pressure of 4.8 bars. Copper rings containing frozen samples were stored in liquid nitrogen until the freeze substitution

procedure was initiated. After high-pressure freezing, samples were transferred to a freeze substitution automate (AFS, Leica Microsystems) pre-cooled to −140°C. As previously described (Ovide et al., 2018), samples were substituted in anhydrous acetone with 0.5% uranyl acetate at −90°C for 96 h. Using a gradient of +2°C h^{-1}, the temperature was gradually raised from −90 to −15°C with two intermediate steps at −60 and −30°C. Finally, samples were rinsed twice with anhydrous ethanol.

Resin infiltration was processed at −15°C in a solution of ethanol/London Resin White (LRW) with successive ratios of 2:1 overday; 1:1 overnight and 1:2 overday followed by a final step in a pure LRW solution renewed twice during 48 h. The LRW was finally polymerized into the AFS apparatus at −15°C under ultra violet light during 48 h. Ultrathin sections (80 nm; ultracut UCT, Leica Microsystems) of diatoms were collected onto carbon-formvar-coated nickel grids. A classical staining using uranyl acetate and lead citrate was done before sections were observed in a Philips, FEI Tecnai 12 Biotwin transmission electron microscope operating at 80 kV, with ES500W Erlangshen CCD camera (Gatan).

Structural Characterization of *P. tricornutum* Morphotypes Through Confocal Microscopy

For confocal microscopy, fluorescent labeling were performed on living or fixed *P. tricornutum* cells. After the different steps of labeling and rinsing, 5 µL of the diatom cell solution were deposited on a 35-mm glass bottom microwell dish (MatTek corporation) and covered with a small agar pad (Fisher, 0.3 g/20 mL) to stabilize microalgae during imaging. Acquisitions were performed at room temperature with an inverted Leica TCS SP5 confocal microscope (Leica Microsystems, Nanterre, France).

Determination of Spectral Characteristics of *P. tricornutum* Cells Autofluorescence Through Confocal Microscopy

One-photon excitation (Ex) and emission (Em) spectra were measured at room temperature using Λλ acquisition mode on a TCS SP5 confocal microscope equipped with a supercontinuum laser source (NTK photonics, Cologne, Germany) and a resonant scanner (8,000 Hz). Using a 63× objective (1.4, oil immersion), autofluorescence emission from diatom cells was detected through a hybrid detector (Leica Microsystems, France). In this configuration, two-dimensional scanning with automatic variations of excitation (Λ, from 470 to 670 nm, 2 nm step) and emission (λ, from 490 to 800 nm, 10 nm band) was performed and led to a stack of 1,722 images (n = 20). Resulting Λλ representation, also called Lambda square fluorescence mapping, was obtained using the Excitation Emission Contour Plot of the Leica Application Suite Advanced Fluorescence software (Leica Microsystems, France). Therefore, each element of the mapping is defined by a corresponding couple of Ex/Em wavelengths. Excitation and emission spectra can therefore be obtained through Microsoft Excel. Consequently, autofluorescence emission of *P. tricornutum* cells was collected from 640 to 720 nm.

Labeling of Living *P. tricornutum* Cells for Nucleic Acids, Mitochondria and Lipid Bodies

To avoid any spectral contamination between cells autofluorescence and green-emitted fluorescent probes for macromolecules and organelles, excitation and emission spectra were measured for Syto 21 and autofluorescence using Λλ acquisition mode as described above and spectral emission windows were determined for each fluorescent component.

For nucleic acids labeling, incubation with Syto 21 (Thermo Fisher Scientific) at a concentration of 10^{-6} M during 5 min, was performed in the respective nutritive medium for fusiform, triradiate, and oval morphotypes. For cell imaging, Syto 21 was excited at 488 nm and fluorescence was collected from 520 to 560 nm. As shown by the Λλ acquisition, activation of the 488 nm wavelength of the supercontinuum laser also induced simultaneous excitation of diatom cells autofluorescence that was detected between 640 and 720 nm. For mitochondria labeling, incubation with Mitotracker Green FM (Thermo Fisher Scientific) at a concentration of 10^{-6} M during 30 min was performed. For cell imaging, Mitotracker green was excited at 488 nm and fluorescence was collected from 500 and 550 nm. For lipid bodies labeling, incubation with BODIPY 505/515 (Thermo Fisher Scientific) at a concentration of 1 mg/mL during 10 min was performed. For cell imaging, BODIPY 505/515 was excited at 500 nm and fluorescence was collected from 510 and 550 nm. As shown by the Λλ acquisition, activation of the 500 nm wavelength of the supercontinuum laser also induced simultaneous excitation of cells autofluorescence that was detected between 640 and 720 nm.

Labeling of Fixed *P. tricornutum* Cells for Actin

Phaeodactylum tricornutum Pt3 cells were fixed in 4% formaldehyde for 1 h and rinsed twice in phosphate buffered saline (PBS). After 3 min pre-incubation with 1% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS, cells were exposed to Alexa Fluor 488-conjugated phalloidin (165 nM, Invitrogen) for 30 min. For cell imaging, Alexa Fluor 488-conjugated phalloidin was excited at 488 nm and fluorescence was collected from 500 and 550 nm.All experiments have been performed at least from three different cell culture and representative images were chosen among at least 20 images to illustrate the different fluorescent labeling.

Relative Quantification of Protein Release From *P. tricornutum* Pt3 Morphotypes

Phaeodactylum tricornutum cells (2.10^5 cells mL^{-1}) were used to inoculate eight flasks for each morphotype, in 100% seawater medium (Instant Ocean) for the triradiate and fusiform morphotypes and 10% for the oval one, respectively. The medium and the culture conditions were as described in the

section "Culture and Growth Conditions of *Phaeodactylum tricornutum*." Each day and for each morphotype, the number of cells was counted in order to establish a growth curve. For each day of culture, culture medium from one flask were recovered by centrifugation at 4,500*g*. Cell pellets were discarded and supernatants containing the culture medium were harvested, dialyzed and lyophilized. The samples were then resuspended in the same volume of milliQ water. To profile the protein release from the different morphotypes, proteins contained in the culture media were separated on a Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). For each morphotype and each day, the volume of secreted medium equivalent to 7.6×10^6 cells were loaded after denaturation using a Laemmli buffer on a SDS-PAGE, ran in a Bis-Tris gel 4-12%. Secreted proteins are finally revealed by silver staining. A 8-bit tiff image of the gel obtained using the Fusion FX6 acquisition system with eVo-6 camera (Vilber). ImageJ (Abràmoff et al., 2004; Rasband, 1997–2018), was used to perform relative quantification of silver-stained proteins. Tiff image was first inverted to finely localize specific staining within an appropriate region of interest (ROI). Subtracted from background noise, the sum of pixel intensity for each ROI was calculated and considered as the indicator of total protein content for each day of culture. All values were normalized by the maximum value detected in the gel i.e., day1 for oval cells and expressed as a kinetic of protein release over days.

Image Analysis

Deconvolution of raw data from confocal imaging was obtained through image processing with Huygens professional 4.5.1 sofware (SVI). ImageJ was used to adjust image brightness and contrast and to perform z projections of 3D images (xyz).

Transcriptome Analysis

The transcriptomic full dataset from Ovide et al., 2018 comparing the oval versus fusiform and the oval versus triradiate cells were combined and were manually reinvestigated in order to identify and select mRNA encoding for fucoxanthin chlorophyll a/c, proteins involved in actin and tubulin network, for signal peptidases and signal recognition particle proteins and finally proteins involved in vesicular trafficking. **Supplementary Tables 1, 2** were build based on these analyses.

RESULTS AND DISCUSSION

The Pt3 strain was adapted to generate enriched cultures in each specific morphotype as previously described in Ovide et al. (2018). Morphotypes were studied and compared with respect to organelles and kinetics of protein secretion.

Ultrastructural Characterization of Pt3 Cells Through Transmission Electron Microscopy

The analysis of the ultrastructure of the three morphotypes was performed by transmission electron microscopy (TEM).

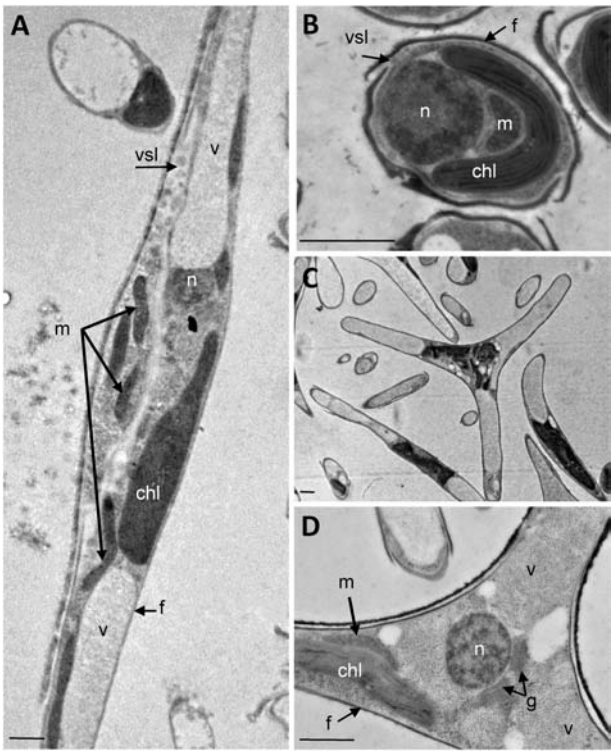


FIGURE 1 | Ultrastructural characterization of *P. tricornutum* Pt3 morphotypes. Transmission Electron micrographs of *P. tricornutum* Pt3 cell morphotypes. Overview of the cells embedded in LRW resin with 0.5% uranyl acetate in a methanol/Reynold's lead citrate solution. **(A)** Fusiform morphotype; **(B)** Oval morphotype, and **(C)** Triradiate morphotype. **(D)** Enlarge view showing cellular distribution organelles in triradiate morphotype. chl: plastid; f: frustule; g: Golgi apparatus; m: mitochondria; n: nucleus; vsl: vesicle; v: vacuole;. Bars, 1 μm.

Electron micrographs of *P. tricornutum* fusiform, oval and triradiate morphotypes are shown in **Figure 1**. As expected, the sections reveal cells surrounded by the frustule which is poorly silicified in the fusiform and triradiate cells (**Figures 1A,C,D**), compared to the oval cells (**Figure 1B**; Borowitzka and Volcani, 1978; Francius et al., 2008; Tanaka et al., 2015). Overall, similar organelles were found in the three morphotypes (**Figure 1**) including nucleus (n), plastid (chl), mitochondria (m), vacuoles (v), vesicles (vsl). Vacuoles are larger in the fusiform and triradiate cells (**Figures 1A,D**). They occupy the distal arms of the cells. A single and large plastid is present and localized nearby the nucleus in the three morphotypes. When observed, Golgi apparatus can be found closed to the nucleus (**Figure 1D**). Mitochondria are elongated and generally reach both extremities of the cells, especially in the fusiform and triradiate cells (**Figures 1A,D**). Such observations correlate with previous description (Martin-Jézéquel and Tesson, 2013) and validate the integrity of *P. tricornutum* cells in the culture conditions used for this work. Then, the three morphotypes of *P. tricornutum* were further characterized with advanced light microscopy by taking advantages of cellular

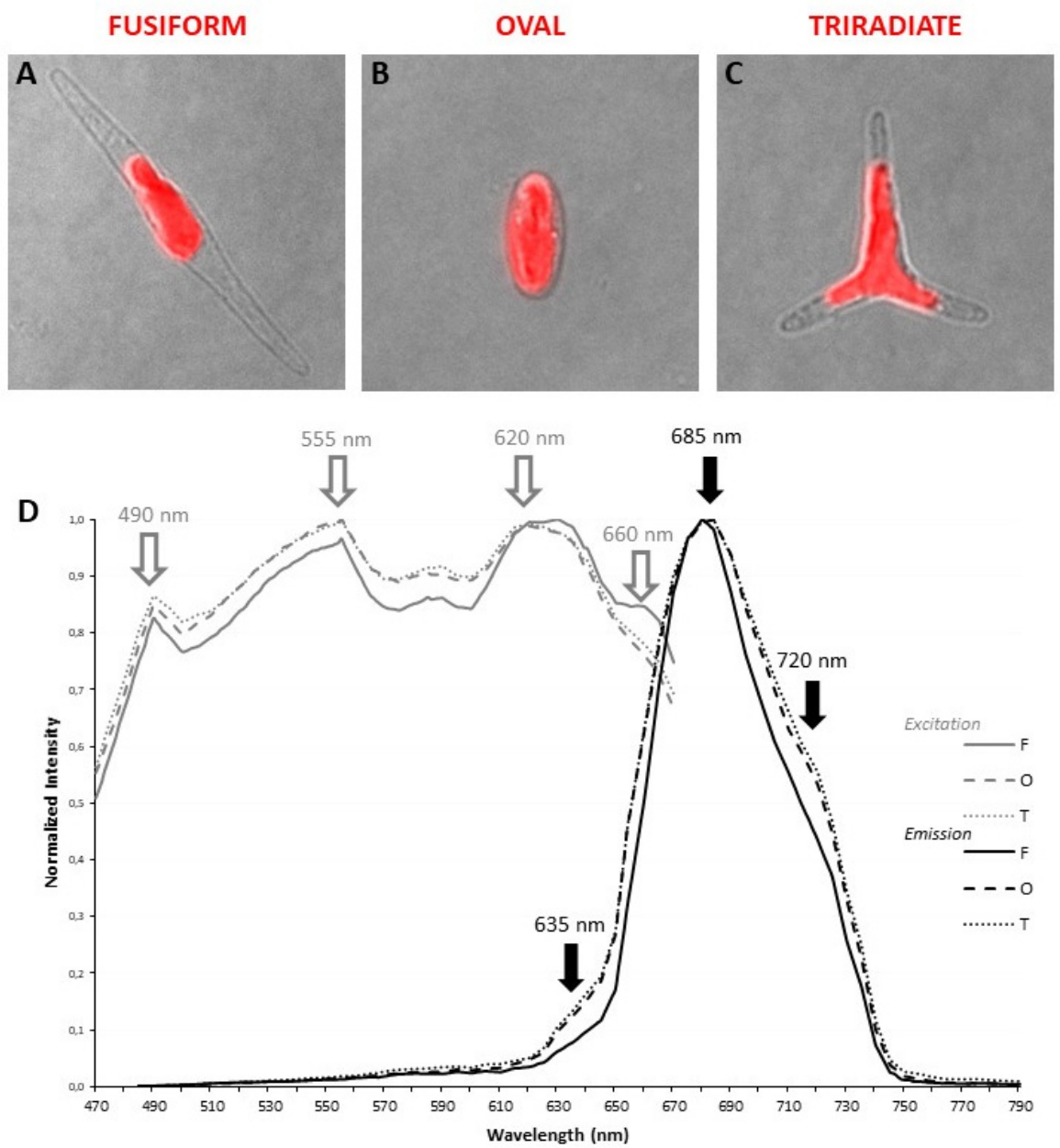


FIGURE 2 | Localization and spectral characterization of plastid autofluorescence in living *P. tricornutum* Pt3 morphotypes through confocal microscopy. Merged images (transmitted light and confocal microscopy) illustrating the shape and the localization of autofluorescent plastid in oval **(A)**, fusiform **(B)**, and triradiate **(C)** morphotype. **(D)** Excitation/emission spectra of plastid autofluorescence were obtained with a supercontinuum laser source through Λλ-scan microscopy (n = 25). Excitation spectra are represented by gray lines. Emission spectra are represented by black lines. Dashed lines: oval cells; solid lines, fusiform cells; dotted lines: triradiate cells.

autofluorescence and labeling of living cells using specific fluorescent organelles probes.

Spectral Characterization of Plastid Autofluorescence in Pt3 Cells

Living *P. tricornutum* cells contain a single and large autofluorescent plastid whose shape is related to the cell organization of each morphotype. In fusiform cells, the organelle is central and presents an elongated aspect (**Figure 2A**) while in triradiate cells, the plastid is located in a central position and is extended in the initial part of the three distal arms (**Figure 2C**). In oval cells, the ovoid plastid occupies a large portion of the cell volume (**Figure 2B**). Thanks to the supercontinuum laser and the Lambda Square mode for signal detection, simultaneous 1P excitation (Λ) and emission (λ)

spectra were obtained for plastid autofluorescence in living cells at room temperature. As described in the Materials and Methods section, the two-dimensional scanning with automatic variations of excitation (Λ, from 470 to 670 nm, 2 nm step) and emission (λ, from 490 to 800 nm, 10 nm band) is particularly appropriate to define robust fluorescence imaging settings for multi-labeling experiments. In contrast, this approach will not allow the discrimination of individual pigments i.e., within excitation profile that requires a single emission wavelength at very low temperature as proposed by Lamote et al. (2003). In the current study, three major peaks of excitation around 490, 555, and 620 nm are observed for all three morphotypes (**Figure 2D**). It is assumed that carotenoids are excited at 490 nm, fucoxanthin at 555 nm while chlorophyll *c* presents a major peak of excitation around 620 nm (Lamote et al., 2003, Veith and Bühel, 2007). In addition, a slight excitation of chlorophyll *a*/*c* with a 490-nm laser cannot completely be ruled out. The fusiform morphotype presents an additional shoulder in its excitation profile above 660 nm. For emission, a major 130 nm-band (full width at half maximum) between 660 and 730 nm with a maximum at 685 nm could be detected and attributed to the light energy collecting complex of photosystem II (PSII) as previously proposed for *Fucus serratus* 8 h-old embryos when excited at 440 nm (Lamote et al., 2003). These results are also consistent with previous data demonstrated that some fucoxanthin chlorophyll *a*/*c* proteins of *P. tricornutum* cells as components of PSII (Levitan et al., 2019) emitted at 683 nm when excited by single laser lines at 473 or 532 nm (Premvardhan et al., 2013). In addition to the major emission peak, two additional shoulders around 630 and 720 nm, respectively, are observed for oval and triradiate morphotypes. The shoulder at 630 nm might reflect that chlorophyll *c* is not integrated within the PSII complex whereas the one at 720 mn suggests that the ratio of PSI/PSII and/or content of lhcf15 could be higher in the oval and triradiate morphotype compare to the fusiform one (Lamote et al., 2003; Herbstová et al., 2017). Whether the latest is related to enhanced energy transfer mechanisms, to an increase in PSI complex or in lhcf15 deserve further investigations. The emission at 720 nm might also be the result of stress conditions (Premvardhan et al., 2013), which is coherent with the fact that oval cells are preponderant under unfavorable growth conditions (De Martino et al., 2007, 2011). *P. tricornutum* genome encodes 42 predicted light-harvesting complex (LHC) or fucoxanthin chlorophyll *a*/*c* proteins (Depauw et al., 2012; Nymark et al., 2013; Levitan et al., 2019).[1] Among them, transcriptomic analysis revealed that some genes like Phatr3_J32294 (lhcr8; UniProt: B7FQS0), Phatr3_J10243 (lhcr9; UniProt: B5Y4K0), Phatr3_J30643 and Phatr3_J29266 (lhcf6; UniProt: B7G5S7), Phatr3_J30031 (lhcf9; UniProt: B7G955), Phatr3_J18049 (lhcf1; UniProt: B7FRW5), and Phatr3_J25172 (lhcf2; UniProt: B7FRW4) are differentially overexpressed in the oval morphotype when compared to the fusiform and triradiate ones (Ovide et al., 2018). Expression of genes like Phatr3_J46529 encoding extrinsic protein in Photosytem II (UniProt: B7G1J1), Phatr3_J11006, and Phatr3_J42519 encoding lhcr1 (UniProt: B7FUM6) and fucoxanthin chlorophyll binding protein related (UniProt: B7FRK1), respectively, are also up-regulated in the oval cells compared to the fusiform and triradiate morphotypes. Accumulation of fucoxanthin had already been described in *P. tricornutum* under low light intensities and depending of the culture conditions (Gómez-Loredo et al., 2016; McClure et al., 2018). Moreover, in diatoms, chlorophyll *a* fluorescence could change as a result of external stimulants or growth phase (Kuczynska et al., 2015) and autofluorescence spectral characteristics can therefore be considered as a "health indicator" during biotechnological applications. Fluorescence Life-time Imaging Microscopy (FLIM) might be further considered to discriminate autofluorescence components and variations as previously proposed by Kuczynska et al. (2015).

Localization of Nucleic Acids Materials in Pt3 Cells

Since *P. tricornutum* plastid autofluorescence exhibits complex spectra for excitation and emission with multiple peaks or shoulders. Simultaneous detection of autofluorescence and other labeling were performed with green light emitting markers only. In addition, similar simultaneous 1P excitation (Λ) and emission (λ) approach was replicated, at least with a DNA/RNA green fluorescent marker named Syto 21 (excitation/emission 494/517 nm), to determine robust spectral configuration for simultaneous fluorescent detection. Similar peaks of excitation at 490, 550, and 620 nm were obtained. Interestingly, the 490 nm-peak, also described as an excitation wavelength for Syto 21, becomes predominant (**Figure 3A**). As expected, autofluorescence emission peaks at 685 and 720 nm were detected but an additional large band of emission from 510 to 600 nm was observed for Syto 21 (**Figure 3A**). Consequently, single excitation at 490 nm and sequential detection between 520–560 nm and 640–720 nm were used for Syto 21 and autofluorescence, respectively. In these conditions, a central rounded nucleus is observed next to the plastid in all three Pt3 morphotypes (**Figures 3B,D,F**) as previously described for other strains (Borowitzka and Volcani, 1978; Siaut et al., 2007; Tanaka et al., 2015; Flori et al., 2017). In addition to nuclear staining, Syto 21-positive materials were also detected around the plastid. In particular, punctiform and sparse Syto 21-positive elements were distributed close to the plastid in oval cells (**Figure 3C**). In fusiform cell, Syto 21-labeling finely delimits the plastid (**Figure 3E**) while staining in triradiate cells was a mix of oval and fusiform ones with both punctiform elements and plastid outlining (**Figure 3G**). Since Syto 21 recognizes both DNA and RNA, extra-nucleus labeling may represent either endoplasmic reticulum (ER) and/or mitochondrial DNA. This is in agreement with the fact that the nuclear envelop has been described to be part of the ER surrounding the plastid (Borowitzka and Volcani, 1978). Moreover, when ER specific proteins like SEC61 subunit or the hDER 1, a central component of the ERAD machinery, were expressed in *P. tricornutum* as eGFP fusion proteins, their localization highlight the ER, the nuclear envelope as well as the outermost membrane of the complex plastid (Liu et al., 2016).

[1] https://www.uniprot.org

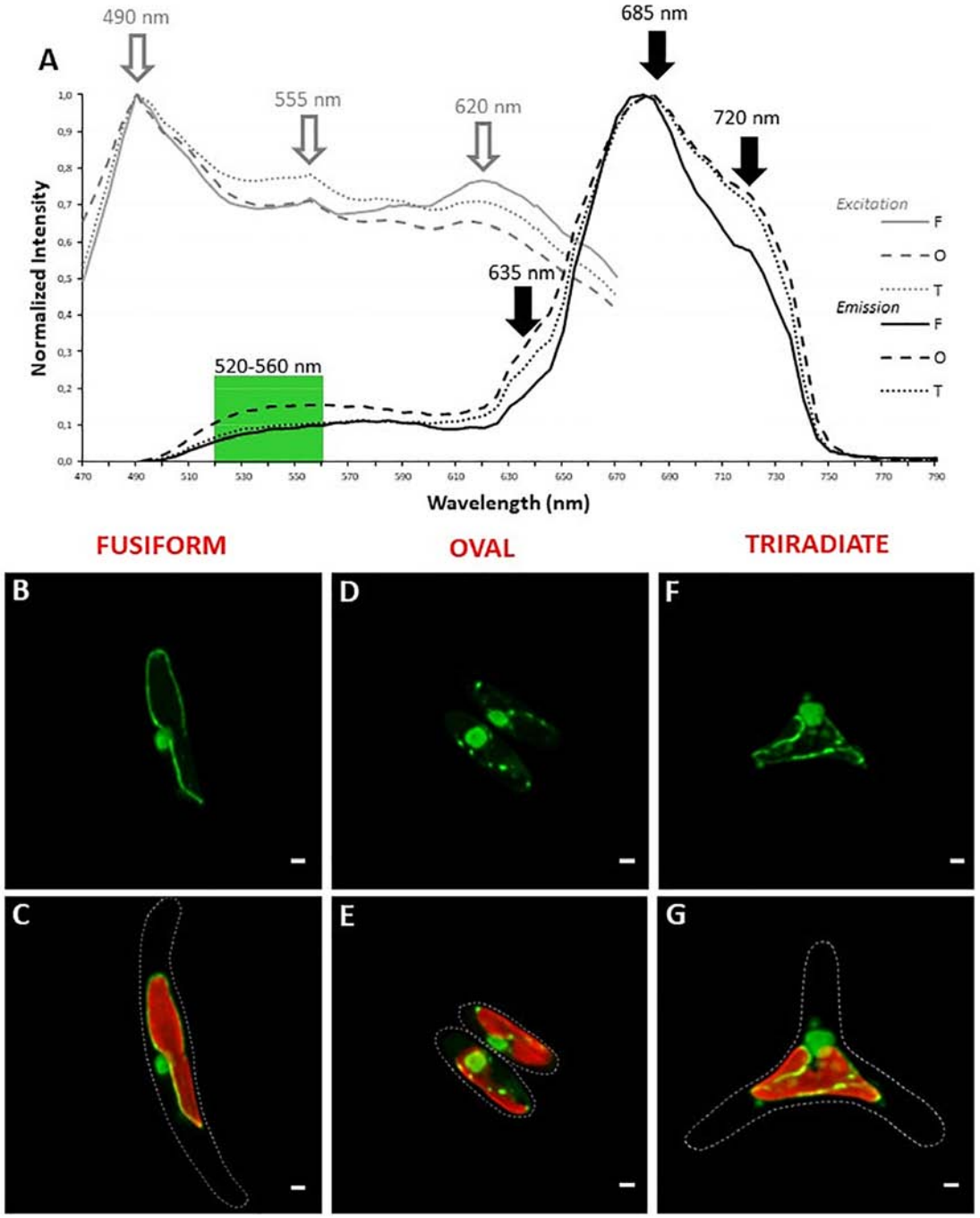


FIGURE 3 | Localization of nucleic acids in living *P. tricornum* Pt3 morphotypes. **(A)** Spectral characterization (Λλ-scan microscopy) for single excitation and double emission settings in Syto 21-labeled and autofluorescent choloroplast-containing cells (n = 20). Excitation spectra are represented by gray lines. Emission spectra are represented by black lines. Dashed lines: oval cells; solid lines, fusiform cells; dotted lines: triradiate cells. **(B,D,F)** Localization of DNA/RNA in Pt3 morphotypes after staining with Syto 21. **(C,E,G)** Combination of Syto 21 images with plastid autofluorescence. Bars, 1 μm.

Localization of Mitochondria in Pt3 Cells

Mitochondria distribution in living *P. tricornutum* cells was studied with the Mitotracker Green probe as previously used for labeling fusiform cells in Liu et al. (2016). In all morphotypes, a moderate to intense Mitotracker Green positive signal delimits the outline of the plastid (**Figure 4A,B,D,E,G,H**). This is in agreement with transmission electron micrographs where tubular mitochondria can be observed close to the plastid in the oval (**Figure 4C**), fusiform (**Figure 4F**), and triradiate cells (**Figure 4I**). Similarly, transgenic *P. tricornutum* Pt1 expressing

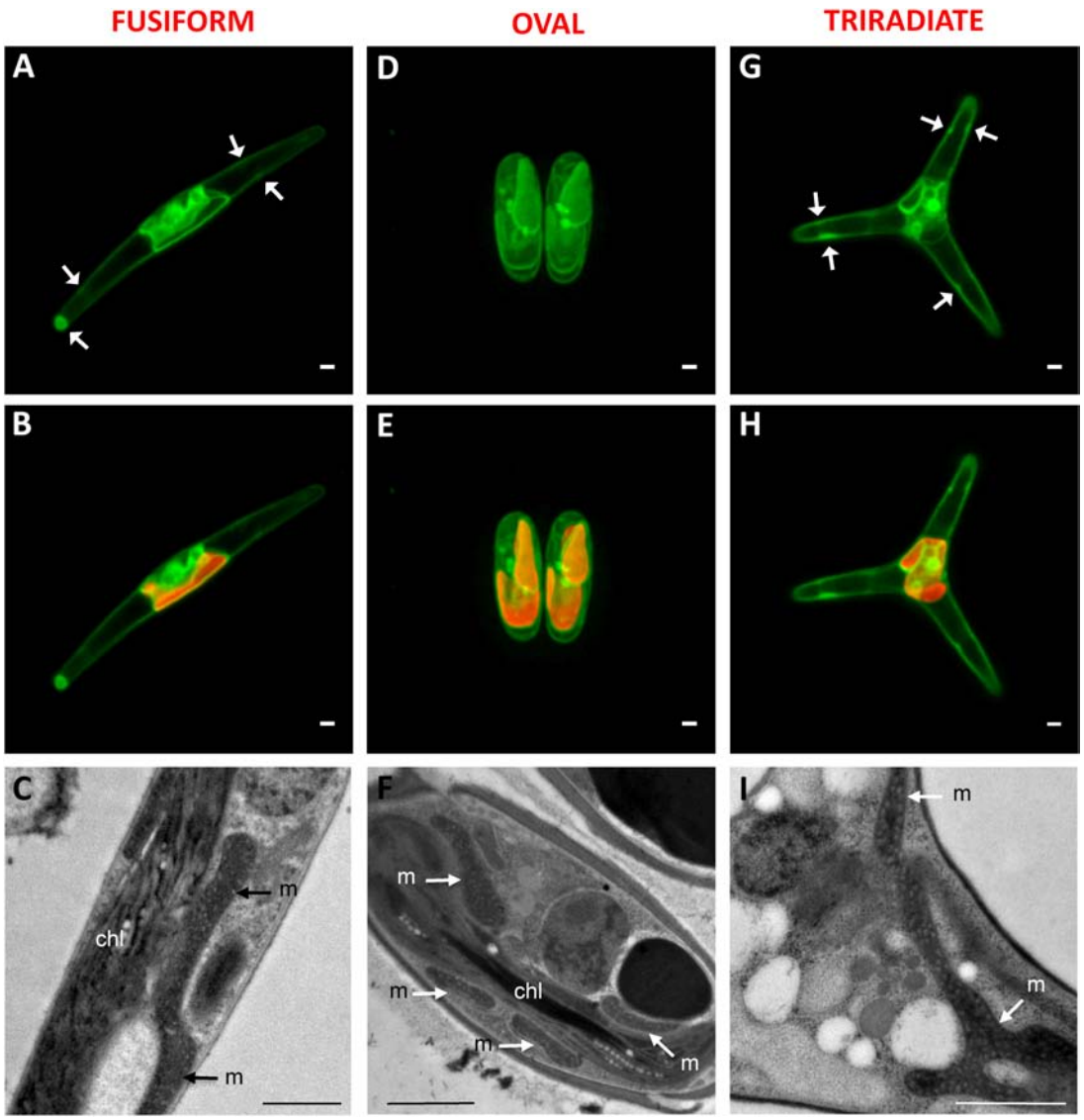


FIGURE 4 | Localization of mitochondria in living *P. tricornutum* Pt3 morphotypes. **(A,D,G)** Localization of mitochondria in Pt3 morphotypes in living cells after staining with Mitotracker Green in the fusiform morphotype **(A)**, oval morphotype **(D)**, and triradiate morphotype **(G)**. White arrows indicate peripheral spots of Mitotracker Green staining. **(B,E,H)** Combination of Mitotracker Green images with plastid autofluorescence in the three Pt3 morphotypes. Enlarged views of tubular mitochondria on Transmission Electron micrograph of each morphotype **(C,F,I)**. chl: plastid; m: mitochondria. Bars, 1 μm.

mitochondrial targeting glutamine synthetase III fusion protein displayed eYFP signal that surrounded the plastid (Siaut et al., 2007). Expression of a mitochondrion marker like a subunit of the glycine decarboxylase complex as an eGFP-fusion protein resulted in a fluorescence pattern near the complex plastid in the fusiform cells (Liu et al., 2016). Moreover, in *P. tricornutum* Pt1 fusiform cells, a continuous network of mitochondria sitting on the plastid is also clearly described through focused ion beam-scanning electron microscopy (Bailleul et al., 2015; Flori et al., 2017, Uwizeye et al., 2020). Such physical contacts between the two organelles may possibly facilitate exchange of energy. In this study, Mitotracker Green-staining was also widely detected in the cytoplasm of the Pt3 cells generally close to the plastid but also within distal arms of fusiform and triradiate morphotypes as peripheral spots (white arrows) (**Figures 4A,G**). Similar elongated branched mitochondrion is also described in fusiform Pt1 cells during interphase (Tanaka et al., 2015) or in tomograms of Pt1 (Uwizeye et al., 2020). Intriguingly, incubation with Mitotracker Green also induced a fluorescence signal along the plasma membrane of oval, fusiform and triradiate cells, suggesting numerous elongated mitochondria at the cell periphery as confirmed by the TEM observation (**Figures 4C,F,I**). Mitotracker Green was chosen in this study for spectral considerations but differential distribution of mitochondria may be noted with different fluorescent Mitotrackers (Galas et al., 2018). In particular, fluorescence of the Mitotracker Orange

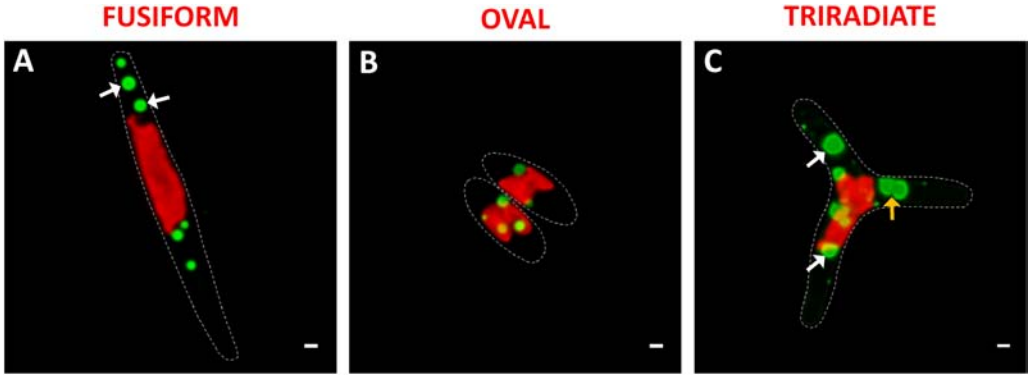


FIGURE 5 | Localization of lipid droplets in living *P. tricornutum* Pt3 morphotypes. **(A–C)** Localization of lipid bodies in Pt3 morphotypes after staining with Bodipy 505/515. Combination of Bodipy 505/515 images with plastid autofluorescence. White arrows indicate lipid bodies with very clear outer ring. Orange arrow illustrate fusion between two lipid bodies. Bars, 1 μm.

and Red probes is dependent of mitochondrial potential while Mitotracker Green is not. Therefore, activity of mitochondria observed close to the plasma membrane as shown in electron micrographs (**Figures 1**, **4**), might be different compared to others observed at the vicinity of plastids. Alternatively, abnormal adsorption of Mitotracker Green on the frustule cannot be completely ruled out since no labeling for nucleic acids belonging to the mitochondrial genome was observed in this area.

Localization of Lipid Bodies in Pt3 Cells

Phaeodactylum tricornutum microalgae synthesize and store neutral lipids mainly triglycerides in lipid bodies also called lipid droplets (Wong and Franz, 2013; Lupette et al., 2019). The distribution of lipid droplets was studied in living *P. tricornutum* Pt3 cells using Bodipy 505/515, which has a small fluorescence Stokes shift and high fluorescence quantum yield for lipids. At this stage of Pt3 culture (day 8), lipid bodies are spherical (**Figure 5**). This contrasts to previous observation of *P. tricornutum* aging culture in which lipid bodies appear as single or double large ovoid lipid droplets (Wong and Franz, 2013). This depends on growth conditions and carbon availability. In this work, small lipid droplets were generally distributed close to the plastid in living oval cells (**Figure 5B**). In contrast, bigger and more numerous lipid bodies were detected in the fusiform and triradiate cells (**Figures 5C,D**). This agrees with a previous report that described lipid droplets in contact with chloroplast (Lupette et al., 2019). The lipid bodies were observed in the distal arms of living fusiform cells as middle size lipid organelles (**Figure 5A**), whereas living triradiate cells contained lipid bodies around the plastid and in distal arms with a large size scale from punctiform to large droplets (**Figure 5C**). In Pt3 cells, lipid bodies are delimited by a ring (white arrows). Previous studies indicate that in *P. tricornutum,* droplets tend to fuse leading to a restricted number of large lipid bodies while in *Tetraselmis suecia* new lipid bodies are synthetized (Wong and Franz, 2013). A possible merge between two lipid droplets is indicated by an orange arrow in **Figure 5C**. From this work, it appears that the Pt3 triradiate cells seems to possess bigger neutral lipids droplets. Interestingly, an additional cruciform morphotype of *P. tricornutum* resulting from triradiate cells transformation with low temperature culture conditions presented a unique fatty acids characteristics suitable for biodiesel production (He et al., 2014). In 2020, Song et al. (2020) observed more and larger lipid bodies in Pt1 and Pt4 fusiform cells over time compared to oval cells. This implies higher neutral lipid accumulation in the fusiform cells from these *P. tricornutum* strains (Song et al., 2020).

Comparison of the Secretory Potential of Pt3 Fusiform, Oval and Triradiate Morphotypes

In diatoms, the molecular mechanisms underlying the release of vesicles including silica deposition vesicles for generation of the silicified frustule (Siaut et al., 2007) or antibody/protein-containing vesicles (Hempel et al., 2011, 2017; Vanier et al., 2015, 2018), are not yet understood and rarely investigated (Erdene-Ochir et al., 2019). In particular, the involvement of cytoskeleton elements in the context of the secretory pathway is poorly described. In this study, the localization of F-actin was determined in fixed permeabilized *P. tricornutum* cells with Alexa 488-Phalloidin. In all three morphotypes, an intense Alexa 488-phalloidin positive signal delimits the outline of the cell indicating the existence of an actin cortex under the plasma membrane (**Figure 6**). Cortical labeling for Syntaxin-A and Sec4 was also observed in Pt1 fusiform cells, suggesting the existence of regulated mechanisms of vesicle fusion and secretion in *P. tricornutum* (Siaut et al., 2007; De Martino et al., 2011). An actin network with a faint fluorescent signal was detected in the cytoplasm of fusiform, oval and triradiate cells (**Figure 6**) suggesting a role of actin in vesicular trafficking together with small GTPase such as SEC4 (Siaut et al., 2007; De Martino et al., 2011). In addition, re-analysis of whole transcriptomic dataset from Ovide et al., 2018 reveals that genes encoding components of the actin network like Phatr3_J44183 encoding the actin cortical patch component lsb4 (UniProt: B5Y5L8) is down-regulated (−1.5 fold) whereas Phatr3_J9601 encoding F-actin capping protein subunit β (UniProt: B7FPL9) and Phatr3_J35252

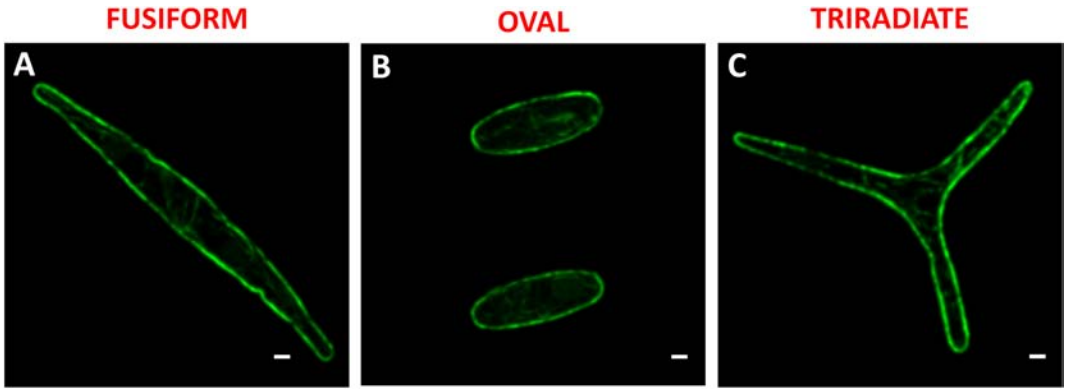


FIGURE 6 | Localization of actin in fixed *P. tricornutum* Pt3 morphotypes. **(A–C)** Localization of F-actin in Pt3 morphotypes after staining with Alexa 488-coupled phalloidin. Bars, 1 μm.

encoding F-actin capping protein (UniProt: B7FXZ8) are up-regulated in the oval morphotype compared to the fusiform and triradiate ones (+1.1 and +1.7 fold, respectively). Other genes encoding for molecular actors associated to actin like Phatr3_J20837 encoding the actin-related protein 4 (no UniProt number available), Phatr3_J48922 encoding condensin complex subunit 3 (UniProt: B7G8V9), genes encoding myosin proteins (Phatr3_EG02335, UniProt: C6JVY2; Phatr3_J52058, UniProt: C6JVY4; Phatr3_J432, UniProt: C6JVY6) are also down-regulated in the oval cells. In contrast, Phatr3_J45476 encoding villin-3-like isoform x1 (UniProt: B7FXU1), Phatr3_J53980 encoding the gelsolin-like protein 2-like (UniProt: B7FPI9), Phatr3_EG02110, UniProt: C6JVY3; Phatr3_EG02422, UniProt: C6JVY5; Phatr3_EG0237 (UniProt: C6JVY7), and Phatr3_J25867 (UniProt: C6JVY8) encoding proteins from the myosin complex and gene encoding the cofilin tropomyosin-type actin-binding protein (Phatr3_EG00210) are up-regulated in the oval cells (Ovide et al., 2018). More experimental work will be needed in the future to decipher the secretion mechanism in *P. tricornutum*.

Despite extensive efforts through immunocytochemistry or silicon rhodamine probe strategies (Galas et al., 2018), we never manage to observe microtubules in *P. tricornutum* as similarly mentioned by Tanaka et al. (2015). This contrast with previous report claiming the observation of microtubules near the nucleus during division of *P. tricornutum* (Borowitzka and Volcani, 1978), *Coscinodiscus granii* and *Entomoneis alata* (Tesson and Hildebrand, 2010). When comparing whole transcriptomes of oval versus fusiform or triradiate cells, only Phatr3_J44333 encoding the gamma-tubulin complex component 3, Phatr3_J17048 encoding the tubulin-specific chaperone a (Uniprot: B7GEH7), and Phatr3_J37751 encoding the tubulin-tyrosine ligase (Uniprot: B7G3N2) have been identified to be slightly overexpressed (between 1.7- and 2.9-fold, respectively) in the oval cells (Ovide et al., 2018). In contrast, many proteins from the kinesin complex and other proteins, which are known to move along or to be associated with the microtubules (**Supplementary Table 1**) are differentially expressed in the oval cells (Ovide et al., 2018). Further investigations need to be performed in order to evaluate whether tubulin components participate to the vesicle trafficking and release in *P. tricornutum*.

So far, *P. tricornutum* cells are known to secrete adhesive mucilage also called exopolymeric substances or EPS. EPS secretion occurs from the GB region in the oval cells. EPS are excreted by the three morphotypes of *P. tricornutum* but significant variations have been observed in the EPS composition, especially between the fusiform and oval cells (Willis et al., 2013). Moreover, secretion of extracellular components like laminarine, elastin, fibronectin, mucin, tenascin have been suggested (Scala et al., 2002; Sapriel et al., 2009). In addition, the capacity of *P. tricornutum* to secrete proteins was highlighted toward proteomic analyses of the culture media. Such analyses allowed the identification of the most abundant proteins, 36 proteins in Buhmann et al. (2016) and 468 proteins in Erdene-Ochir et al. (2019), respectively. Among the secreted proteins in the culture medium of *P. tricornutum* cells, the "highly abundant secreted protein 1" (HASP1; Uniprot: B7G4A0) also called phosphate alkaline was confirmed by LC-MS/MS (Buhmann et al., 2016; Erdene-Ochir et al., 2019). The HASP1 signal peptide drives the entry of protein into the secretory pathway (Erdene-Ochir et al., 2019). Moreover, when the recombinant mAb directed against the hepatitis B was expressed in *P. tricornutum* cells, the human signal peptide from both the heavy and light chains were cleaved off in the diatom cells suggesting that they used a signal peptide peptidase mechanism that is similar to the one occurring in other eukaryotes (Vanier et al., 2015). All the past studies were performed on the fusiform cells independently of the morphotype of the *P. tricornutum* cells. In this work, we checked on previous transcriptomic dataset regarding the comparative mRNA expression between oval and fusiform and triradiate morphotypes, respectively by looking at the putative signal peptidases involved in the removal of the signal peptides (Ovide et al., 2018). This search allows the identification of Phatr3_J18533, Phatr3_J51280, Phatr3_J15399 and Phatr3_J13921 genes encoding respectively four signal peptidases (Uniprot: B5Y4T0; B7GDX7, B7G8T6, B7G343) that are overexpressed between 2.6- to 4.2-fold when comparing the oval versus fusiform and triradiate cells, thus

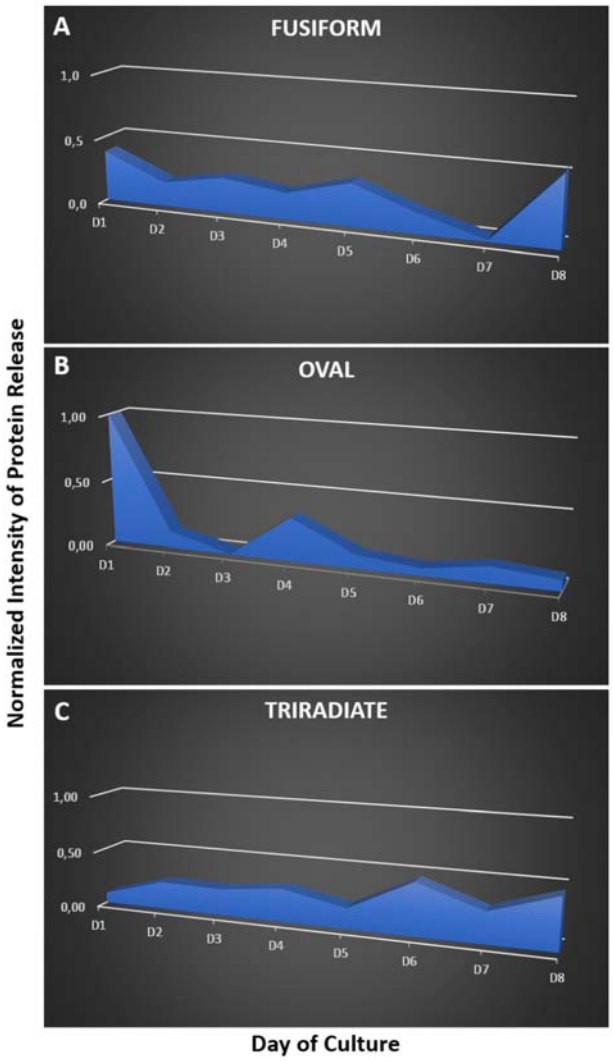


FIGURE 7 | Kinetics of protein release by the three morphotypes of *P. tricornutum* Pt3 strain over a 8 days period of culture. The proteins present in the culture medium of fusiform **(A)**, oval **(B)**, and triradiate **(C)** were separated by SDS-PAGE. Proteins in the gel were labeled with silver staining. Relative quantification of stained proteins were performed with ImageJ.

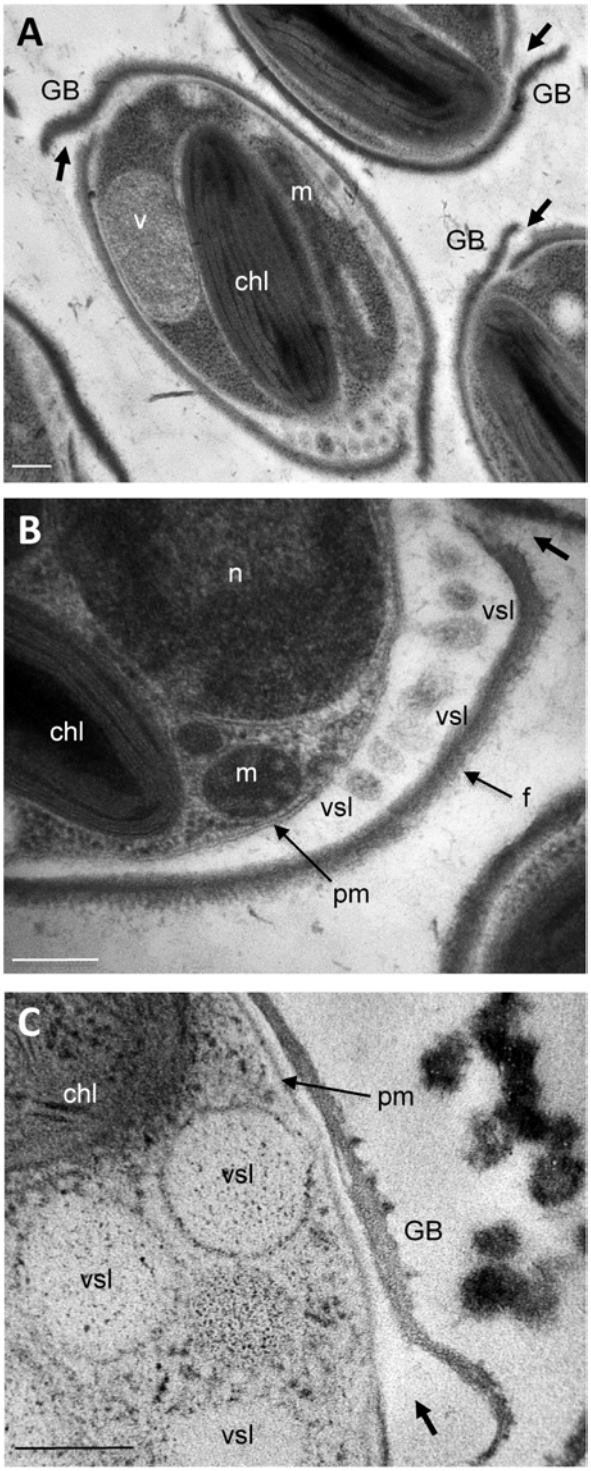


FIGURE 8 | Ultrastructural characterization of the release site of secretory vesicles in *P. tricornutum* oval morphotype. Transmission Electron micrographs of Pt3 *Phaeodactylum tricornutum* oval morphotype illustrating the secretion through vesicles. **(A)** Whole view of the oval cell. **(B)** Magnification in the interspace between the plasma membrane and the frustule where secretory vesicles are accumulated. **(C)** Secretory vesicles nearby the girdle band. chl: plastid; GB: girdle band; m: mitochondria; n: nucleus; pm: plasma membrane; f: frustule; vsl: secretory vesicles v: vacuole. Bars, 0.2 μm.

suggesting that the oval cells present higher secretion capacity. Phatr3_J44439 gene encoding the signal recognition particle 19kDa protein (Uniprot: B7FU41) is also overexpressed 3.5 fold in the oval cells as compared to the fusiform and triradiate cells. In contrast, signal recognition particle proteins like the one encoded by Phatr3_EG02041 gene, the SRP54 encoded by Phatr3_J13417 gene (Uniprot: B5Y444) and the signal recognition particle 72 kDa protein encoded by Phatr3_J48508 gene (Uniprot: B7G7I1) are slightly down-regulated (less than 2 fold; Ovide et al., 2018). In addition, Phatr3_J47612 gene encoding the HASP1 protein (Uniprot: B7G4A0) is up regulated more than 4 fold in the oval cells compared to the fusiform and triradiate cells. In this context, in order to compare the kinetics of protein secretion of the three morphotypes, the

spontaneous protein release in the culture medium of each morphotype was followed over a culture period of 8 days using SDS-PAGE (**Figure 7** and **Supplementary Figure 1**). Such analysis highlights different kinetics of protein release between Pt3 cells. Indeed, the fusiform cells secreted proteins rather constantly (**Figure 7A**). In contrast, within the first 4 days, oval cells released more than 80% of their proteins while triradiate cells secreted only 38% (**Figures 7B,C**). In addition, triradiate cells increased regularly their secretory activity over the 8-days period (**Figure 7C**). Taken together, these data suggest differences in secretory kinetics between Pt3 cells. Oval cells are able to release rapidly a higher amount of proteins, fusiform cells present a constant secretory activity at a mid-level while triradiate cells release progressively proteins over time from low-level to mid-level. Such results are complementary to recent findings published by Song et al. (2020) that quantified higher protein content in oval cell cultures than in fusiform cell cultures for both Pt1 and Pt4 strains of *P. tricornutum*. In addition, as previously reported in Ovide et al. (2018), RNA-Seq transcriptomic analysis performed on the three morphotypes of *P. tricornutum* Pt3 strain highlighted, in the oval morphotype, overexpression of genes encoding proteins involved in vesicular transports like the SAR1, a GTPase found in COP II vesicles; BET1 a Golgi vesicular transport from the ER to the Golgi complex; the SNARE SEC22 and the syntaxin 6, which displays important role in protein trafficking between the *trans*-Golgi network and the endosomal system. In this study, other genes encoding proteins involved in vesicular trafficking are observed to be up and down-regulated in the oval morphotype (**Supplementary Table 2**). This includes genes encoding Clathrin-heavy chain (Phatr3_EG01984, UniProt: B7G4Y3), COP I (Phatr3_J49956, UniProt: B7GCF6) and COP II (Phatr3_J49955, UniProt: B7GCF; Phatr3_J47710, UniProt: B7G4M2) that are down-regulated. Several genes encoding ARF and RAB-related proteins have been identified to be down- or up-regulated suggesting a fine-tune regulation of the secretion in the oval cells (**Supplementary Table 2**). Interestingly, genes encoding coatomers like Phatr3_J19093 gene that encodes the coatomer subunit epsilon (Uniprot: B7FUJ7) and Phatr3_J7018 gene that encode the coatomer subunit zeta-1 (Uniprot: B7G7H3) are up-regulated in the oval cells. These coatomers might be involved in the retrograde vesicle-mediated transport from Golgi apparatus to ER. Such results might be confirmed experimentally in future studies.

As far as we know from the literature, the mechanisms of protein release in *P. tricornutum* cells has not been studied in details yet. Imaging oval cell at subcellular levels revealed frustule opening valves (thick arrow) that might be involved in the secretory process (**Figure 8A**). In particular, a number of electron-dense vesicles are accumulated in the interspace between the plasma membrane and the frustule (**Figure 8B**). Despite reduced cell volume, oval cells seem to be very active concerning secretion (**Figure 8C**). In the future, functional studies will be necessary to depict the complete transport mechanisms of proteins in the three morphotypes of *P. tricornutum* and the dynamics of protein release. This is of particular interest as *P. tricornutum* has been used recently to produce recombinant monoclonal antibodies. However, the production yield is insufficient to envision any industrial commercialization. Thus, gaining comprehension of *P. tricornutum* cellular and metabolism processes would be helpful in the future to maximize the use of *P. tricornutum* as a green alternative cell biofactory. The results presented in this study suggests that using oval cells for the production of biopharmaceutical proteins might be helpful to improve the production yield. In addition, characterizing the secretory pathways by which proteins such as recombinant mAbs are released would be of particular interest in this blue biotech context.

CONCLUSION

The work described herein revealed characteristics of cellular organelles, cytoskeleton and protein secretion in the three main morphotypes of *P. tricornutum* Pt3 fusiform, oval and triradiate.

The three morphotypes share similarities including spectral characteristics of the plastid, the location of the nucleus, the organization of mitochondria around the plastid as well as the existence of both a F-actin cortex and an intracellular network of F-actin. In contrast, the oval cell, which is the smallest Pt3 morphotype, presents a thick frustule and a plastid occupying a large cell volume. As compare to fusiform and triradiate cells, oval cells release spontaneously proteins more rapidly. In addition, comparison of whole transcriptomes of oval versus fusiform or triradiate cells revealed numerous differential expression of positive and negative regulators belonging to the complex dynamic secretory machinery. Since such processes are mostly regulated at the protein level, future proteomic analyses is required to gain informations regarding the fine regulation of secretion in the three Pt3 morphotypes.

This study highlights the specificities occurring within the oval morphotype confirming that the oval cells secrete more rapidly proteins. Thus, in the future, using oval cells for the production of biopharmaceutical proteins might be helpful to improve the production yield. Indeed, due to easy transformation procedure, *P. tricornutum* fusiform cells are currently used to produce recombinant mAbs directed against viruses. However, the production yield of the secreted recombinant mAbs is still low (2 mg L^{-1}) and insufficient to envision an industrial commercialization (Hempel and Maier, 2012). In addition, characterizing the secretory pathway(s) by which proteins such as recombinant mAbs are released would be of particular interest and will help maximizing the future use of *P. tricornutum* as a green alternative cell biofactory.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

LG, MBa, and CB: concept and design of the research and writing of the manuscript. LG, CB, DS, SB, and MBé: experimental work. LG, MBa, CB, MR, DS, and MBé: data analysis and interpretation. All authors have read, corrected, and agreed on the content of the manuscript prior to its submission.

FUNDING

The authors are grateful for financial support from the IUF (Institut Universitaire de France Junior 2014-2019); the European Union's Horizon 2020 Research and Innovation programme under the Grant Agreement 774078 (Pharma-Factory); the European Regional Development Fund (ERDF—PACT-CBS); the Region Normandie through the Normandy Plant Technologies (NPT) project; and the University of Rouen Normandie (UNIROUEN), INSERM, IRIB, Normandie University, IBiSA.

ACKNOWLEDGMENTS

We thank Prof. J.-C. Mollet and Dr. I. Boulogne, coordinators of the Normandy Plant Technologies (NPT) project financed by the Region Normandie for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.638181/full#supplementary-material

REFERENCES

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with ImageJ. *Biophot. Intern.* 11, 36–43.

Apt, K. E., Grossman, A. R., and Kroth-Pancic, P. G. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Molec. Gen. Genet.* 252, 572–579.

Baïet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M.-C., et al. (2011). N-glycans of *Phaeodactylum tricornutum* diatom and functional characterization of its N-acetylglucosaminyltransferase I enzyme. *J. Biol. Chem.* 286, 6152–6164. doi: 10.1074/jbc.m110.175711

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., Tanaka, A., et al. (2015). Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* 524, 366–369. doi: 10.1038/nature14599

Bartual, A., Gálvez, J. A., and Ojeda, F. (2008). Phenotypic response of the diatom *Phaeodactylum tricornutum* Bohlin to experimental changes in the inorganic carbon system. *Bot. Mar.* 51, 350–359.

Bezanilla, M. (2013). What can plants do for cell biology? *Mol. Biol. Cell.* 24, 2491–2493. doi: 10.1091/mbc.e12-10-0706

Borowitzka, M. A., and Volcani, B. E. (1978). Polymorphic diatom *Phaeodactylum tricornutum*: ultrastructure of its morphotypes. *J. Phycol.* 14, 10–21. doi: 10.1111/j.1529-8817.1978.tb00625.x

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244.

Bowler, C., Vardi, A., and Allen, A. E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Annu. Rev. Mar. Sci.* 2, 333–365. doi: 10.1146/annurev-marine-120308-081051

Buhmann, M. T., Schulze, B., Förderer, A., Schleheck, D., and Kroth, P. G. (2016). Bacteria may induce the secretion of mucin-like proteins by the diatom *Phaeodactylum tricornutum*. *J. Phycol.* 52, 463–474. doi: 10.1111/jpy.12409

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). *Phaedactylum tricornutum*: a diatom cell factory. *Trends Biotechnol.* 38, 606–622. doi: 10.1016/j.tibtech.2019.12.023

Cadoret, J.-P., Garnier, M., and Saint-Jean, B. (2012). Microalgae, functional genomics and biotechnology. *Adv. Bot. Res.* 64, 285–341. doi: 10.1007/978-3-540-74335-4_17

Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* 5:3831.

D'Adamo, S., Schiano di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A., et al. (2019). Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* 17, 75–87. doi: 10.1111/pbi.12948

De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., et al. (2011). Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*. *Protist* 162, 462–481. doi: 10.1016/j.protis.2011.02.002

De Martino, A., Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (*Bacillariophyceae*). accessions. *J. Phycol.* 43, 992–1009. doi: 10.1111/j.1529-8817.2007.00384.x

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom *Phaeodactylum tricornutum*. *Nucleic Acids Res.* 37:e96. doi: 10.1093/nar/gkp448

Depauw, F., Rogato, A., D'Alcalá, M., and Falciatore, A. (2012). Exploring the molecular basis of responses to light in marine diatoms. *J. Exper. Bot.* 63, 1575–1591. doi: 10.1093/jxb/ers005

Dumontier, R., Mareck, A., Mati-Baouche, N., Lerouge, P., and Bardor, M. (2018). "Toward future engineering of the N-glycosylation pathways in microalgae for optimizing the production of biopharmaceuticals," in *Microalgal Biotechnology*, eds E. Jacob-Lopes, L. Q. Zepka, and M. I. Queiroz (London: IntechOpen), doi: 10.5772/intechopen.73401

Erdene-Ochir, E., Shin, B. K., Kwon, B., Jung, C., and Pan, C. H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Sci Rep.* 9:9941. doi: 10.1038/s41598-019-45786-9

Flori, S., Jouneau, P. H., Bailleul, B., Gallet, B., Estrozi, L. F., Moriscot, C., et al. (2017). Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nat. Commun.* 8:15885. doi: 10.1038/ncomms15885

Francius, G., Tesson, B., Dague, E., Martin-Jézéquel, V., and Dufrêne, Y. F. (2008). Nanostructure and nanomechanics of live *Phaeodactylum tricornutum* morphotypes. *Environ. Microbiol.* 10, 1344–1356. doi: 10.1111/j.1462-2920.2007.01551.x

Galas, L., Gallavardin, T., Bénard, M., Lehner, A., Schapman, D., Lebon, A., et al. (2018). "Probe, Sample, and Instrument (PSI)": the hat-trick for fluorescence live cell imaging. *Chemosensors* 6:40. doi: 10.3390/chemosensors6030040

Gómez-Loredo, A., Benavides, J., and Rito-Palomares, M. (2016). Growth kinetics and fucoxanthin production of *Phaeodactylum tricornutum* and *Isochrysis galbana* cultures at different light and agitation conditions. *J. Appl. Phycol.* 28, 849–860. doi: 10.1007/s10811-015-0635-0

Gutenbrunner, S. A., Thalhamer, J., and Schmid, A.-M. M. (1994). Proteinaceous and immunochemical distinctions between the oval and fusiform morphotypes of *Phaeodactylum tricornutum* (*Bacillariophyceae*). *J. Phycol.* 30, 129–136. doi: 10.1111/j.0022-3646.1994.00129.x

Hamilton, M. L., Warwick, J., Terry, A., Allen, M. J., Napier, J. A., and Sayanova, O. (2015). Towards the industrial production of omega-3 long chain polyunsaturated fatty acids from a genetically modified diatom *Phaeodactylum tricornutum*. *PLoS One* 10:e0144054. doi: 10.1371/journal.pone.0144054

He, L., Han, X., and Yu, Z. (2014). A Rare *Phaeodactylum tricornutum* cruciform morphotype: culture conditions, transformation and unique fatty acid characteristics. *PLoS One* 9:e93922. doi: 10.1371/journal.pone.0093922

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One* 6:e28424. doi: 10.1371/journal.pone.0028424

Hempel, F., and Maier, U. G. (2012). An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microb. Cell. Fact.* 11:126. doi: 10.1186/1475-2859-11-126

Hempel, F., and Maier, U. G. (2016). Microalgae as solar-powered protein factories. *Adv. Exp. Med. Biol.* 896, 241–262. doi: 10.1007/978-3-319-27216-0_16

Hempel, F., Maurer, M., Brockmann, B., Mayer, C., Biedenkopf, N., Kelterbaum, A., et al. (2017). From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein. *Microb. Cell. Fact.* 16:131. doi: 10.1186/s12934-017-0745-2

Herbstová, M., Bína, D., Kaòa, R., Vácha, F., and Litvín, R. (2017). Red-light phenotype in a marine diatom involves a specialized oligomeric red-shifted antenna and altered cell morphology. *Sci. Rep.* 7:11976. doi: 10.1038/s41598-017-12247-0

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925. doi: 10.1038/ncomms7925

Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618. doi: 10.1038/nrg2386

Kuczynska, P., Jemiola-Rzeminska, M., and Strzalka, K. (2015). Photosynthetic pigments in diatoms. *Mar. Drugs* 13, 5847–5881. doi: 10.3390/md13095847

Lamote, M., Darko, E., Schoefs, B., and Lemoine, Y. (2003). Assembly of the photosynthetic apparatus in embryos from *Fucus serratus* L. *Photosynth. Res.* 77, 45–52. doi: 10.1023/A:1024999024157

León-Bañares, R., González-Ballester, D., Galván, A., and Fernández, E. (2004). Transgenic microalgae as green cell-factories. *Trends Biotechnol.* 22, 45–52. doi: 10.1016/j.tibtech.2003.11.003

Levitan, O., Chen, M., Kuang, X., Cheong, K. Y., Jiang, J., Banal, M., et al. (2019). Structural and functional analyses of photosystem II in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 17316–17322. doi: 10.1073/pnas.1906726116

Lewin, J. C., Lewin, R. A., and Philpott, D. E. (1958). Observations on *Phaeodactylum tricornutum*. *J. Gen. Microbiol.* 18, 418–426.

Liu, X. J., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., et al. (2016). Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Res.* 20, 249–257. doi: 10.1016/j.algal.2016.10.018

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., Schaller, H., et al. (2019). The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*. *Algal Res.* 38:101415. doi: 10.1016/j.algal.2019.101415

Mann, M., Serif, M., Jakob, T., Kroth, P. G., and Wilhelm, C. (2017). PtAUREO1a and PtAUREO1b knockout mutants of the diatom *Phaeodactylum tricornutum* are blocked in photoacclimation to blue light. *J. Plant Physiol.* 217, 44–48. doi: 10.1016/j.jplph.2017.05.020

Martin, S. G. (2014). Yeasts as models in cell biology. *FEMS Microbiol. Rev.* 38:143. doi: 10.1111/1574-6976.12068

Martin-Jézéquel, V., and Tesson, B. (2013). "*Phaeodactylum tricornutum* polymorphism: an overview," in *Advances in Algal Cell Biology*, eds K. Heimann and C. Katsaros (Berlin: De Gruyter), 43–80.

Mathieu-Rivet, E., Kiefer-Meyer, M. C., Vanier, G., Ovide, C., Burel, C., Lerouge, P., et al. (2014). Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* 5:359. doi: 10.3389/fpls.2014.00359

Mathur, M., Xiang, J. S., and Smolke, C. D. (2017). Mammalian synthetic biology for studying the cell. *J. Cell. Biol.* 216, 73–82. doi: 10.1083/jcb.201611002

McClure, D. D., Luiz, A., Gerber, B., Barton, G. W., and Kavanagh, J. M. (2018). An investigation into the effect of culture conditions on fucoxanthin production using the marine microalgae *Phaeodactylum tricornutum*. *Algal Res.* 29, 41–48. doi: 10.1016/j.algal.2017.11.015

Miyahara, M., Aoi, M., Inoue-Kashino, N., Kashino, Y., and Ifuku, K. (2013). Highly efficient transformation of the diatom *Phaeodactylum tricornutum* by multi-pulse electroporation. *Biosci. Biotechnol. Biochem.* 77, 874–876. doi: 10.1271/bbb.120936

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324, 1724–1726. doi: 10.1126/science.1172983

Niu, Y. F., Yang, Z. K., Zhang, M. H., Zhu, C. C., Yang, W. D., Liu, J. S., et al. (2012). Transformation of diatom *Phaeodactylum tricornutum* by electroporation and establishment of inducible selection marker. *Biotechniques* 52, 1–3.

Nymark, M., Sharma, A., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951.

Nymark, M., Valle, K. C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., et al. (2013). Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to Re-illumination in the diatom *Phaeodactylum tricornutum*. *PLoS One* 8:e58722. doi: 10.1371/journal.pone.0058722

Ovide, C., Kiefer-Meyer, M. C., Bérard, C., Vergne, N., Lecroq, T., Plasson, C., et al. (2018). Comparative in depth RNA sequencing of *P. tricornutum*'s morphotypes reveals specific features of the oval morphotype. *Sci. Rep.* 8:14340. doi: 10.1038/s41598-018-32519-7

Premvardhan, L., Réfrégiers, M., and Büchel, C. (2013). Pigment organization effects on energy transfer and Chl a emission imaged in the diatoms *C. meneghiniana* and *P. tricornutum in vivo*: a confocal laser scanning fluorescence (CLSF). microscopy and spectroscopy study. *J. Phys. Chem.* 117, 11272–11281. doi: 10.1021/jp402094c

Prihoda, J., Tanaka, A., de Paula, W. B. M., Allen, J. F., Tirichine, L., and Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* 63, 1543–1557.

Rasband, W. S. (1997–2018). *ImageJ*. Bethesda, MA: U.S. National Institutes of Health. Available online at: https://imagej.nih.gov/ij/

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., 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. doi: 10.1038/s41598-018-23106-x

Rosales-Mendoza, S., Solís-Andrade, K. I., Márquez-Escobar, V. A., González-Ortega, O., and Bañuelos-Hernandez, B. (2020). Current advances in the algae-made biopharmaceuticals field. *Expert Opin. Biol.* 20, 751–766. doi: 10.1080/14712598.2020.1739643

Sapriel, G., Quinet, M., Heijde, M., Jourdren, L., Tanty, V., Luo, G., et al. (2009). Genome-wide transcriptome analyses of silicon metabolism in *Phaeodactylum tricornutum* reveal the multilevel regulation of silicic acid transporters. *PLoS One* 4:e7458. doi: 10.1371/journal.pone.0007458

Sasso, S., Pohnert, G., Lohr, M., Mittag, M., and Hertweck, C. (2012). Microalgae in the postgenomic era: a blooming reservoir for new natural products. *FEMS Microbiol. Rev.* 36, 761–785. doi: 10.1111/j.1574-6976.2011.00304.x

Scala, S., Carels, N., Falciatore, A., Chiusane, M. L., and Bowler, C. (2002). Genome properties of the diatom *P. tricornutum*. *Plant Physiol.* 129, 996–1002.

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9:3924.

Serif, M., Lepetit, B., Weissert, K., Kroth, P. G., and Rio Bartulos, C. (2017). A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*. *Algal Res.* 23, 186–195. doi: 10.1016/j.algal.2017.02.005

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A., et al. (2007). Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. *Gene* 406, 23–35. doi: 10.1016/j.gene.2007.05.022

Slattery, S. S., Diamond, A., Wang, H., Therrien, J. A., Lant, J. T., Jazey, T., et al. (2018). An expanded plasmid-based genetic toolbox enables cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS Synth. Biol.* 7, 328–338. doi: 10.1021/acssynbio.7b00191

Song, Z., Lye, G. J., and Parker, B. M. (2020). Morphological and biochemical changes in *Phaedactylum tricornutum* triggered by culture media: Implications for industrial exploitation. *Algal Res.* 47:101822. doi: 10.1016/j.algal.2020.101822

Stukenberg, D., Zauner, S., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 9:740. doi: 10.3389/fpls.2018.00740

Tanaka, A., De Martino, A., Amato, A., Montsant, A., Mathieu, B., Rostaing, P., et al. (2015). Ultrastructure and membrane traffic during cell division in the marine pennate diatom *Phaeodactylum tricornutum*. *Protist* 166, 506–521. doi: 10.1016/j.protis.2015.07.005

Tesson, B., and Hildebrand, M. (2010). Extensive and intimate association of the cytoskeleton with forming silica in diatoms: control over patterning on the meso- and micro-scale. *PLoS One* 10:e14300. doi: 10.1371/journal.pone.0014300

Uwizeye, C., Decelle, J., Jouneau, P.-H., Gallet, B., Keck, J.-B., Moriscot, C., et al. (2020). In-cell quantitative structural imaging of phytoplankton using 3D electron microscopy. *bioRxiv* [Preprint], doi: 10.1101/2020.05.19.104166

Vanier, G., Hempel, F., Chan, P., Rodamer, M., Vaudry, D., Maier, U. G., et al. (2015). Biochemical characterization of human anti-Hepatitis B monoclonal antibody produced in the microalgae *Phaeodactylum tricornutum*. *PLoS One* 10:e0139282. doi: 10.1371/journal.pone.0139282

Vanier, G., Stelter, S., Vanier, J., Hempel, F., Maier, U. G., Lerouge, P., et al. (2018). Alga-made anti-Hepatitis B antibody binds to human Fcγ receptors. *Biotechnol. J.* 13:e1700496. doi: 10.1002/biot.201700496

Vartanian, M., Desclés, J., Quinet, M., Douady, S., and Lopez, P. J. (2009). Plasticity and robustness of pattern formation in the model diatom *Phaeodactylum tricornutum*. *New Phytol.* 182, 429–442. doi: 10.1111/j.1469-8137.2009.02769.x

Veith, T., and Bühel, C. (2007). The monomeric photosystem I-complex of the diatom *Phaeodactylum tricornutum* binds specific fucoxanthin chlorophyll proteins (FCPs) as light-harvesting complexes. *Biochim. Biophys. Acta* 1767, 1428–1435. doi: 10.1016/j.bbabio.2007.09.004

Willis, A., Chiovitti, A., Dugdale, T. M., and Wetherbee, R. (2013). Characterization of the extracellular matrix of *Phaeodactylum tricornutum* (*bacillariophyceae*): Structure, composition, and adhesive characteristics. *J. Phycol.* 49, 937–949. doi: 10.1111/jpy.12103

Wilson, D. P. (1946). The triradiate and other forms of *Nitschia closterum* (Ehrenberg). Wm. Smith form *Minutissima* of Allen and Nelson. *J. Mar. Biol. Assoc.* 26, 235–270. doi: 10.1017/s002531540001211x

Wong, D. N., and Franz, A. K. (2013). A comparison of lipid storage in *Phaeodactylum tricornutum* and *Tetraselmis suecica* using laser scanning confocal microscopy. *J. Microbiol. Methods* 95, 122–128. doi: 10.1016/j.mimet.2013.07.026

Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum* (*Bacillariophyceae*). with a variety of selectable marker and reporter genes. *J. Phycol.* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x

Zhang, C., and Hu, H. (2014). High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Mar. Genomics* 16, 63–66. doi: 10.1016/j.margen.2013.10.003

Zou, L., Chen, J., Zheng, D., Balamurugan, S., Li, D. W., Yang, W. D., et al. (2018). High-efficiency promoter-driven coordinated regulation of multiple metabolic nodes elevates lipid accumulation in the model microalga *Phaeodactylum tricornutum*. *Microb. Cell. Fact.* 17:54. doi: 10.1186/s12934-018-0906-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Galas, Burel, Schapman, Ropitaux, Bernard, Bénard and Bardor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 April 2021
doi: 10.3389/fpls.2021.657821

Carbon Dioxide Concentration Mechanisms in Natural Populations of Marine Diatoms: Insights From *Tara* Oceans

Juan José Pierella Karlusich[1,2], *Chris Bowler*[1,2] *and Haimanti Biswas*[3]*

[1] *Institut de Biologie de l'ENS, Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, France,* [2] *CNRS Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, Paris, France,* [3] *CSIR National Institute of Oceanography, Biological Oceanography Division, Dona Paula, India*

Marine diatoms, the most successful photoautotrophs in the ocean, efficiently sequester a significant part of atmospheric CO_2 to the ocean interior through their participation in the biological carbon pump. However, it is poorly understood how marine diatoms fix such a considerable amount of CO_2, which is vital information toward modeling their response to future CO_2 levels. The *Tara* Oceans expeditions generated molecular data coupled with *in situ* biogeochemical measurements across the main ocean regions, and thus provides a framework to compare diatom genetic and transcriptional flexibility under natural CO_2 variability. The current study investigates the interlink between the environmental variability of CO_2 and other physicochemical parameters with the gene and transcript copy numbers of five key enzymes of diatom CO_2 concentration mechanisms (CCMs): Rubisco activase and carbonic anhydrase (CA) as part of the physical pathway, together with phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase, and malic enzyme as part of the potential C4 biochemical pathway. Toward this aim, we mined >200 metagenomes and >220 metatranscriptomes generated from samples of the surface layer of 66 globally distributed sampling sites and corresponding to the four main size fractions in which diatoms can be found: 0.8–5 μm, 5–20 μm, 20–180 μm, and 180–2,000 μm. Our analyses revealed that the transcripts for the enzymes of the putative C4 biochemical CCM did not in general display co-occurring profiles. The transcripts for CAs were the most abundant, with an order of magnitude higher values than the other enzymes, thus implying the importance of physical CCMs in diatom natural communities. Among the different classes of this enzyme, the most prevalent was the recently characterized iota class. Consequently, very little information is available from natural diatom assemblages about the distribution of this class. Biogeographic distributions for all the enzymes show different abundance hotspots according to the size fraction, pointing to the influence of cell size and aggregation in CCMs. Environmental correlations showed a complex pattern of responses to CO_2 levels, total phytoplankton biomass, temperature, and nutrient concentrations. In conclusion, we propose that biophysical CCMs are prevalent in natural diatom communities.

Keywords: ***Tara*** **Oceans, diatoms, carbon metabolism, carbon dioxide concentration mechanisms, metagenomics, metatranscriptomics**

OPEN ACCESS

Edited by:
Benoit Schoefs,
Le Mans Université, France

Reviewed by:
Ansgar Gruber,
Institute of Parasitology, Academy of Sciences of the Czech Republic (ASCR), Czechia
Yusuke Matsuda,
Kwansei Gakuin University, Japan

***Correspondence:**
Haimanti Biswas
haimanti.biswas@nio.org

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *24 January 2021*
Accepted: *23 March 2021*
Published: *30 April 2021*

Citation:
Pierella Karlusich JJ, Bowler C and Biswas H (2021) Carbon Dioxide Concentration Mechanisms in Natural Populations of Marine Diatoms: Insights From Tara Oceans. Front. Plant Sci. 12:657821. doi: 10.3389/fpls.2021.657821

INTRODUCTION

Diatoms are among the most successful and diversified eukaryotic photoautotrophs in the present day ocean (Armbrust, 2009; Malviya et al., 2016; Pierella Karlusich et al., 2020). Their fast growth rates in high-nutrient environments and comparatively large sizes make them important contributors to organic carbon production. On an annual scale, marine diatoms fix 10–20 billion metric tons of inorganic carbon (comparable to all global rainforests combined), corresponding to up to 40% of the total marine primary production and as much as 20% of the total primary production on Earth (Field et al., 1998; Smetacek, 1999; Granum et al., 2005; Jin et al., 2006; Falkowski and Raven, 2013; Tréguer et al., 2018). Thus, diatoms are main contributors to marine food chains and in sequestering atmospheric CO_2 to the ocean interior through gravitational sinking of particles (biological carbon pump) (**Figure 1A**), and hence have high biogeochemical significance (Tréguer et al., 2018; Boyd et al., 2019). Diatoms possess a peculiar gene complement derived from green and red algal sources, and have many genes in common with animals and bacteria (Bowler et al., 2008; Moustafa et al., 2009; Dorrell et al., 2017, 2021), mostly owing to their chimeric evolutionary origins as well as to horizontal gene transfer events (Armbrust, 2009; Dorrell et al., 2021). It is believed that these genes have enabled them to develop unique and highly efficient carbon (Schoefs et al., 2017) and nitrogen metabolism pathways (Wilhelm et al., 2006; Busseni et al., 2019). In the present context of climate change and the substantial anthropogenic perturbations in the ocean (increasing CO_2 and temperature, acidification, disturbances in nutrient cycles, etc.), a key question is how marine diatoms will respond. In order to do this, a clear understanding of diatom carbon metabolism is required.

The key carbon fixing enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase, Rubisco, is one of the most abundant proteins on Earth and is responsible for 100 billion tons of carbon fixation annually (Erb and Zarzycki, 2018; Bar-On and Milo, 2019). Remarkably however, Rubisco is highly inefficient because it shows specificity toward O_2 which competes with CO_2. This is believed to be a remnant of its evolution at a time when oxygen levels were minimal, and leads to a wasteful process called photorespiration (Poudel et al., 2020). Diatoms possess a red algal type Rubisco (type ID) which is one of the most efficient Rubisco forms, with the highest preference toward carboxylation over oxygenation (Young et al., 2016). However, CO_2 concentration in the present day surface oceans is on average 10–12 μmol kg^{-1} [<1% of the dissolved inorganic carbon (DIC) pool] (**Figure 1A**), which is well below the half saturation constant of Rubisco for CO_2 (Badger et al., 1998). Instead, the DIC pool is >90% bicarbonate ions (HCO_3^-) (**Figure 1A**). To optimize carboxylation even at present CO_2 levels, most photoautotrophs have developed active carbon dioxide concentration mechanisms (CCMs) which can be biophysical or biochemical (**Figure 1B**). Such CCMs aim to maintain a higher CO_2 concentration over O_2 in the vicinity of Rubisco (Reinfelder, 2011). In a biophysical CCM, the cells actively pump bicarbonate ions (HCO_3^-) inside the cell followed by the conversion of HCO_3^- to CO_2 (Hopkinson et al., 2011; Reinfelder, 2011) via the metalloenzyme carbonic anhydrase (CA; Morel et al., 1994; Badger, 2003). CO_2 molecules enter the cell through the lipid bilayer membrane (Gutknecht et al., 1977) and can easily diffuse out due to the high concentration gradient. It has been proposed that in diatoms cytoplasmic CA continuously maintains low CO_2 levels by converting CO_2 to HCO_3^- for facilitating a CO_2 diffusive influx (Matsuda et al., 2017). Hence, CA plays a key role in carbon acquisition in diatoms. The functioning of biophysical CCMs in marine diatoms has been well studied in laboratory conditions (Hopkinson et al., 2016; reviewed by Matsuda et al., 2017) and was shown to be highly diverse and more efficient than in C4 plants (Young et al., 2016). Down regulation of CCM/photorespiratory genes under elevated CO_2 levels in model marine diatoms has been observed in experimental studies (Ohno et al., 2012; Hennon et al., 2015; Li et al., 2015).

In biochemical CCMs, the enzyme phosphoenolpyruvate carboxylase (PEPC) works as a primary carboxylase in the cytoplasm, forming oxaloacetate (C4) from phosphoenolpyruvate (C3) and HCO_3^- (**Figure 1B**). This C4 acid is then transported into the chloroplast and releases CO_2 in the vicinity of Rubisco by action of the enzyme phosphoenolpyruvate carboxykinase (PEPCK) (Reinfelder et al., 2000, 2004; Roberts et al., 2007a,b; and references therein). The process of decarboxylation can also be performed by the malic enzyme (ME). Oxaloacetate (OAA) is converted to malate via malate dehydrogenase which is then transferred to another compartment (likely mitochondria) and forms pyruvate and CO_2 via ME (Kustka et al., 2014). Co-occurrences of PEPCK and ME driven decarboxylation mechanisms have been reported in C4 plants (Cacefo et al., 2019) and marine diatoms (Kroth et al., 2008). It has been proposed that ME may not be actively involved in CCMs and probably plays a role in photorespiration and mitochondrial metabolism in marine diatoms (Davis et al., 2017). The study by Kroth et al. (2008) stated that in case of a C4 pathway in diatoms, the processes of decarboxylation of OAA as well as malate and carboxylation by Rubisco may take place separately in mitochondria and plastids, respectively. In such a case the CO_2 molecule released in mitochondria via decarboxylation needs to be transferred to Rubisco. It is possible that the CO_2 is then converted to HCO_3^- again via CA and a further conversion to CO_2 takes place within the plastid in the vicinity of Rubisco before carboxylation. These double conversions involve a considerable amount of energy and diatoms may use a C4 CCM for dissipating extra energy which they acquire via the light reactions. Thus, diatoms living under optimum light conditions might actively use a C4 CCM, whereas the diatoms from light limited areas and in deep chlorophyll maxima may down-regulate this process to avoid energy loss.

However, the existence of a fully functional biochemical CCM (C4 pathway) in marine diatoms is not yet proven (Tanaka et al., 2014; Clement et al., 2016) despite some experimental studies (Roberts et al., 2007b; Kustka et al., 2014). A short term metabolic C^{14} labeling study of two model marine diatoms (*Thalassiosira pseudonana* and *Thalassiosira weissflogii*) showed that the initial labeled products in *T. pseudonana* were mostly C3 and C6, whereas *T. weissflogii* produced a mixture of C3 and C4

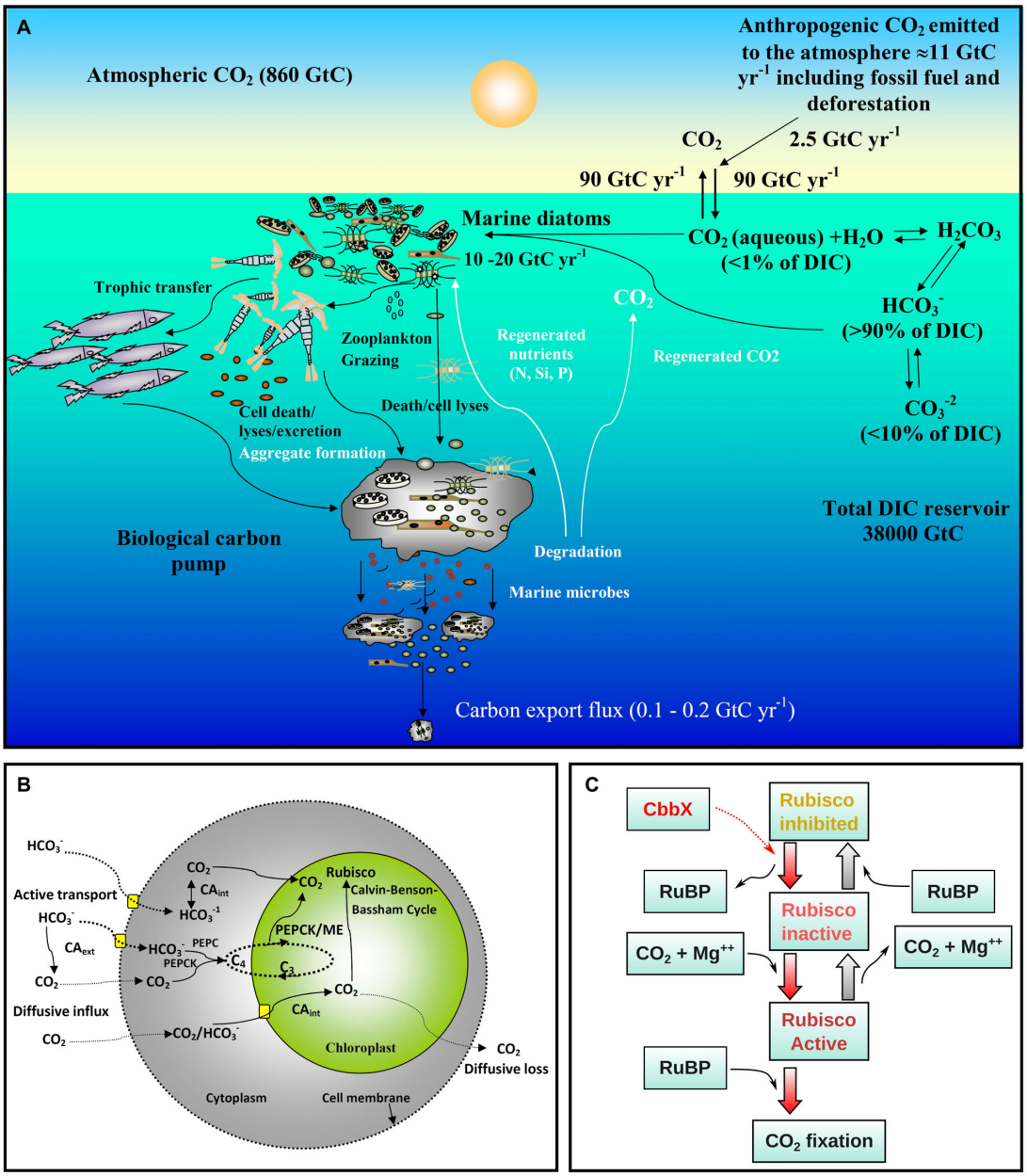


FIGURE 1 | Overview of ocean carbon cycle and diatom carbon dioxide concentration mechanisms. **(A)** Schematic representation of the ocean carbon cycle depicting the role of marine diatoms in the biological carbon pump. The anthropogenic CO_2 emission to the atmosphere (mainly generated by fossil fuel burning and deforestation) is nearly 11 Gigaton carbon (GtC) per year, of which almost 2.5 GtC is taken up by the surface ocean. In surface seawater (pH 8.1–8.4), bicarbonate (HCO_3^-) and carbonate ions (CO_3^{2-}) constitute nearly 90 and <10% of dissolved inorganic carbon (DIC) respectively, while dissolved CO_2 (CO_2 aqueous) contributes <1%. Despite this low level of CO_2 in the ocean and its slow diffusion rate in water, diatoms fix 10–20 GtC annually via photosynthesis thanks to their carbon dioxide concentration mechanisms (CCMs), allowing them to sustain food chains. In addition, 0.1–1% of this organic material produced in the euphotic layer sinks down as particles, thus transferring the surface carbon toward the deep ocean and sequestering atmospheric CO_2 for thousands of years or longer. The remaining organic matter is remineralized through respiration. Thus, diatoms are one of the main players in this biological carbon pump, which is arguably the most important biological mechanism in the Earth System allowing CO_2 to be removed from the carbon cycle for very long period. Based on data from Friedlingstein et al., 2020. **(B)** Schematic representation of the CCMs in diatoms. The low levels of CO_2 in the ocean and its slow diffusion rate in water have led diatoms and other photosynthetic organisms to evolve CCMs that utilize the higher concentrations of HCO_3^-. The biophysical CCM consists of various bicarbonate transporters and carbonic anhydrases (CAs) that serve to increase the CO_2 flux balance toward the pyrenoid, a low CO_2-permeable subcellular compartment in the chloroplast containing most of the Rubisco. In addition, some diatoms may also have a biochemical (C4-like) CCM involving phosphoenolpyruvate carboxylase (PEPC), phosphoenolpyruvate carboxykinase (PEPCK) and/or malic enzyme (ME). **(C)** Schematic presentation of Rubisco activation by CbbX in diatoms and other phototrophs with red-type Rubisco. CbbX functions as a mechanochemical motor protein and uses the energy from ATP hydrolysis to modify the structure of Rubisco. This process facilitates the dissociation of inhibitory sugar phosphates [ribulose-1,5-bisphosphate (RuBP) and others] from the active site of Rubisco.

acids (Roberts et al., 2007b). Notwithstanding, C4 enzymes were documented in both species. This suggests that some diatoms may operate a mixture of C3 and C4 CCMs. A significant increase in expression of genes encoding C4 enzymes under low CO_2 acclimatized cells was reported in model marine diatoms (Kroth et al., 2008; Saade and Bowler, 2009). However, the evidence for an active biochemical CCM in natural communities of marine diatoms has remained inconclusive.

Another inefficient feature of Rubisco in green algae and land plants is its deactivation by sugar phosphates (ribulose-1,5-bisphosphate and others). To perform optimum photosynthesis, Rubisco is usually reactivated by a motor protein, named Rubisco activase (RCA), by binding to the inactive Rubisco via ATP hydrolysis (Shivhare and Mueller-Cajar, 2017; **Figure 1C**). The gradient of pH and Mg^{++} concentrations are two key factors that control RCA activity. A non-substrate CO_2 and a Mg^{++} ion need to bind to Rubisco before carboxylation and therefore the concentration of CO_2 is also important for activation of Rubisco prior to carboxylation (Pollock et al., 2003). In the study by Young et al. (2016) it was noticed that the activation levels of Rubisco in eleven experimental diatom species were quite low suggesting a strong possibility for the presence of a RCA type of enzyme. Surprisingly, no structural homolog of RCA has been reported in diatoms. Instead, a functional homolog of RCA, denoted Calvin-Benson-Bassham protein (CbbX) complex, was identified from a red type Rubisco in proteobacteria (Mueller-Cajar et al., 2011) and red algae (Loganathan et al., 2016). Jensen et al. (2017) reported a BLAST search that revealed the presence of CbbX homologs in almost 100 stramenopiles including some model diatoms. The authors also established that CbbX is encoded in the plastid genome unlike in green plants where the *RCA* gene is encoded in the nucleus. However, it was subsequently found that in red algae and diatoms, another CbbX gene is also encoded in the nucleus (Bhat et al., 2017). Jensen et al. (2017) also argued in favor of the existence of an allosteric control of Rubisco by CbbX in diatoms (**Figure 1C**). Other than this, to our knowledge there have been no other studies of the abundance and functioning of CbbX in diatoms, neither in lab studies nor in natural populations. Conversely, only a few discrete studies have reported gene expression within phytoplankton communities as a function of changing ocean carbon chemistry (Endo et al., 2015; Hennon et al., 2015; Hopkinson et al., 2016). Moreover, most of the studies are based on model diatoms and hence there is a strong need to study natural diatom assemblages.

Therefore, we deemed it important to characterize the diatom CCM in the environment under natural CO_2 variability. With this motivation, the present study investigates the interlink between the abundance and expression of the genes encoding five key enzymes (CbbX, CA, PEPC, PEPCK, and ME) involved in diatom CCMs under variable CO_2 levels. We did so by mining the *Tara* Oceans datasets (**Figure 2**), which were generated from samples across the global ocean in a standardized manner, including the measurement of carbonate chemistry and other physicochemical parameters and the generation of >200 metagenomes and >220 metatranscriptomes (Carradec et al., 2018).

MATERIALS AND METHODS

Sequence Search and Analysis in the *Tara* Oceans Eukaryotic Gene Catalog

We searched for sequences of interest in version 1 of the Marine Atlas of *Tara* Oceans Unigenes (MATOU.v1; Carradec et al., 2018). It consists of 116 million transcribed sequences mainly from eukaryotic plankton in size fractions ranging from 0.8 to 2,000 μm. It was generated by assembling 441 poly-A + metatranscriptomes from samples across the main ocean basins (with the exception of the Arctic Ocean) and then clustered at 95% identity to define a non-redundant catalog (Carradec et al., 2018).

A HMMer search (version 3.2.1 with gathering threshold option)[1] was performed in the translated version of MATOU.v1 using the following Pfam models: PF00004 (AAA; ATPase family associated with various cellular activities) for detecting CbbX, PF00311 (PEPcase; Phosphoenolpyruvate carboxylase) for PEPC, PF01293 (PEPCK_ATP; Phosphoenolpyruvate carboxykinase) for PEPCK, PF00390 (malic; Malic enzyme N-terminal domain) and PF03949 (Malic_M; Malic enzyme NAD binding domain) for ME, PF00194 (Carb_anhydrase; Eukaryotic-type carbonic anhydrase) for alpha-CA, PF00484 (Pro_CA; Carbonic anhydrase) for beta-CA, PF00132 (Hexapep; Bacterial transferase hexapeptide) for gamma-CA, PF10563 (CA_like; Putative carbonic anhydrase) for delta-CA, PF18484 (CDCA; Cadmium carbonic anhydrase repeat) for zeta-CA, PF18599 (LCIB_C_CA; Limiting CO_2-inducible proteins B/C beta carbonic anhydrases) for theta-CA, and PF08332 (CaMKII_AD; Calcium/calmodulin dependent protein kinase II association domain) for iota-CA. To compare with primary and housekeeping pathways, we also retrieved the sequences coding for the nuclear-encoded subunits of photosystem II (PF05151, PsbM; PF01716, MSP; PF05757, PsbQ; PF06514, PsbU; PF18240, PSII_Pbs31) and for ribosomal proteins (112 Pfam models listed in **Supplementary Table 1**). Taxonomic assignment of MATOU.v1 is already available based on sequence similarity against a reference database containing UniRef90, MMETSP, and other sources (see Carradec et al., 2018). Based on this assignment, we only kept with those sequences assigned as diatoms for further analysis.

In order to discard homologous proteins of interest, we carried out a combination of sequence similarity network and phylogeny approaches for functional assignment. Briefly, we carried out a HMMer v3.2.1 search (as previously mentioned) for sequences containing the Pfam domains of interest among the sequenced genomes available in the Integrated Microbial Genome (IMG) database[2] (Chen et al., 2018) and the sequenced transcriptomes from MMETSP (Keeling et al., 2014). The retrieved sequences were translated in the correct frame and the Pfam domain region was extracted. These sequences were used for building a protein similarity network using EFI-EST tool (Zallot et al., 2019) and Cytoscape visualization (Shannon et al., 2003), which allowed us

[1] http://hmmer.org/

[2] http://img.jgi.doe.gov

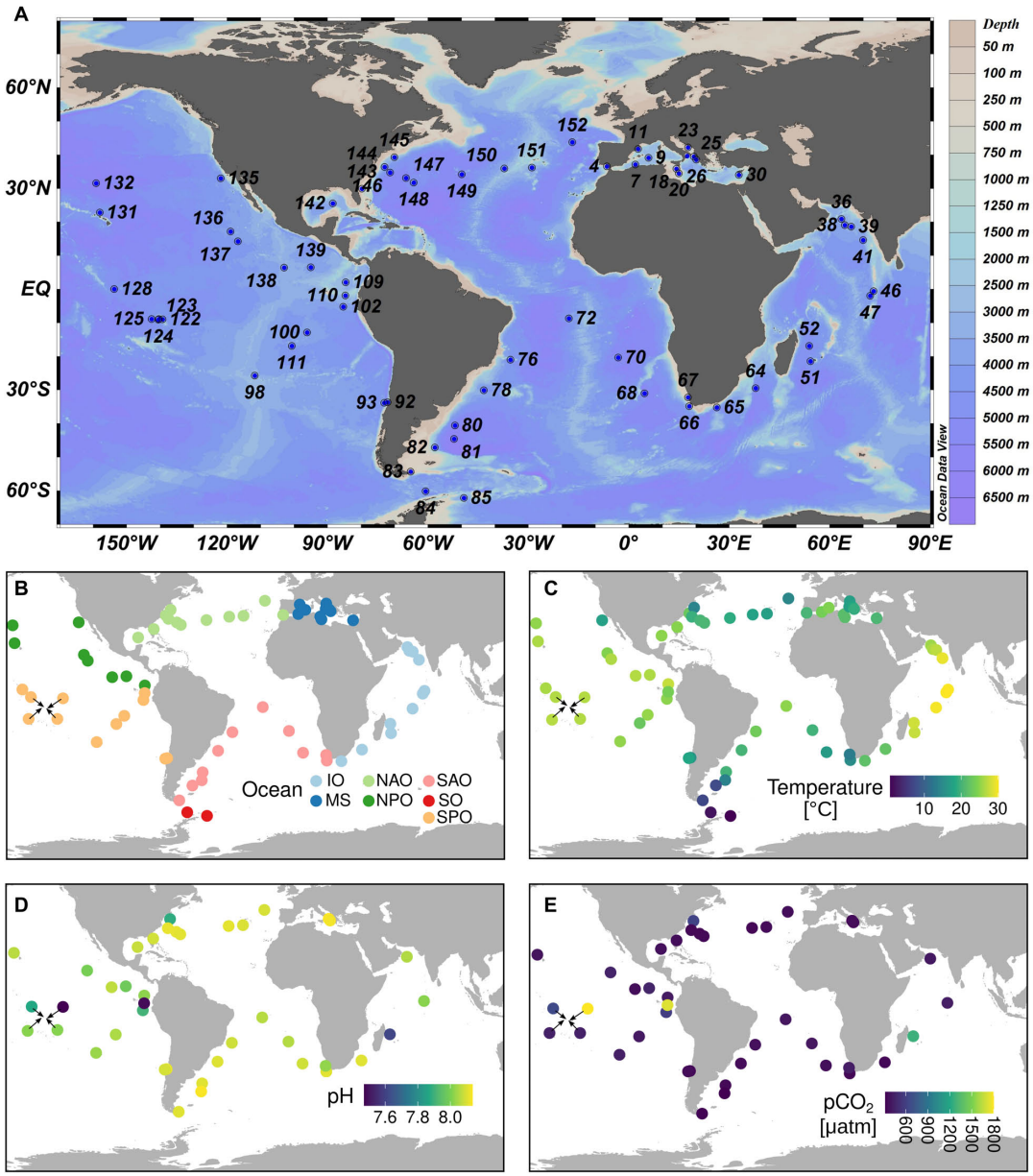


FIGURE 2 | *Tara* Oceans sampling sites relevant to the current study. **(A)** Station labels. **(B)** Ocean regions. **(C)** Temperature measurements. **(D)** pH measurements. **(E)** CO_2 partial pressure measurements. The sampling covers almost all main ecogeographic locations. Complete contextual data is available in **Supplementary Table 1**. IO, Indian Ocean; MS, Mediterranean Sea; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean.

to inspect the different protein clusters varying the score cut-off. By this step, we found that most Pfams were specific to the enzymes of interest (at least in diatoms) with the only exceptions of CbbX and iota CA (**Supplementary Figures 1**, **2**). The final list of MATOU.v1 sequences used in the current work is displayed in **Supplementary Table 1**.

In the case of CbbX, it is part of one of the many clusters detected in the sequence similarity network of the AAA domain sequences (**Supplementary Figure 1A**). This network was built using a score cut-off of 40 after a previous step of reducing sequence redundancy to 80% identity with CD-HIT version 4.6.4 (Li and Godzik, 2006). Therefore, we then built a phylogeny for all the AAA domain sequences of this cluster (**Supplementary Figures 1B,C**). For this, we aligned the sequences with MAFFT version 6 using the G-INS-I strategy (Katoh and Toh, 2008) and used the resulting alignment to generate the tree with PhyML version 3.0 (Guindon et al., 2010). Four categories of rate variation were used. The starting tree was a BIONJ tree and the type of tree improvement was subtree pruning and regrafting. Branch support was calculated using the approximate likelihood

ratio test (aLRT) with a Shimodaira–Hasegawa-like (SH-like) procedure. CbbX sequences formed a distinctive branch (**Supplementary Figure 1B**), which included the experimentally validated sequences from the proteobacterium *Rhodobacter sphaeroides* and the nuclear- and plastid-encoded versions from the red alga *Cyanidioschyzon merolae* (Loganathan et al., 2016). The remaining branches of the tree are annotated as stage V sporulation protein K (KEEG id: K06413) by BlastKOALA (Kanehisa et al., 2016). Therefore, the sequence similarity network and the phylogenies were used as references for the selection of *Tara* Oceans unigenes coding for diatom CbbX.

In the case of iota-CA, it forms one of the two main clusters in the sequence similarity network of CaMKII_AD domain sequences (**Supplementary Figure 2**), which was built using a score cut-off of 18 and a previous step of reducing redundancy at 90% identity with CD-HIT version 4.6.4 (Li and Godzik, 2006). The iota-CA cluster contains sequences from bacteria and eukaryotes, including the experimentally validated iota-CA from *T. pseudonana* (Jensen et al., 2019) as well as orthologous sequences from other species (Jensen et al., 2019; Nonoyama et al., 2019). The other subfamily contains eukaryotic sequences annotated as canonical Calcium/calmodulin dependent protein kinases. Therefore, we used the protein similarity network to keep exclusively with the iota-CAs among those MATOU-v1 sequences with the CaMKII_AD domain.

Analysis of Biogeographical and Environmental Patterns of Gene and Transcript Abundances

Tara Oceans performed a worldwide sampling of plankton between 2009 and 2013 (**Figure 2** and **Supplementary Table 2**) using a serial filtration system for separating the plankton into discrete size fractions (Pesant et al., 2015). In the current work, we analyzed a total of 203 metagenomes and 224 metatranscriptomes generated from samples of the surface layer (5 m depth) of 66 globally distributed stations (**Figure 2**) and corresponding to the four main size fractions enriched in protists: 0.8–5 μm, 5–20 μm, 20–180 μm, and 180–2,000 μm (Carradec et al., 2018). Thus, we retrieved the metagenomic and metatranscriptomic read abundances of the selected MATOU.v1 sequences (described in the section "Sequence Search and Analysis in the *Tara* Oceans Eukaryotic Gene Catalog") and normalized them by the total read abundance for genes or transcripts of the whole diatom community of the corresponding sample. Results are displayed in **Supplementary Table 3**.

We compared the metagenomic and metatranscriptomic abundance patterns with the environmental data collected during *Tara* Oceans expeditions.[3] The contextual data used in the current work is displayed in **Supplementary Table 2**. Carbonate chemistry was determined in 40 stations.[4] Total alkalinity and DIC were measured potentiometrically (Edmond, 1970), and other carbonate chemistry parameters (pH on total scale, CO_2, pCO_2, HCO_3^-, CO_3^-) were calculated using seacarb (Nisumaa et al., 2010). The average CO_2 values were $12 \pm 2 \mu mol\ kg^{-1}$ which is very common for present day surface seawater values. However, there were four stations (TARA_110, TARA_122, TARA_052, TARA_145) with more than double these CO_2 levels. The station locations were highly diverse; from tropical, subtropical and higher latitude locations (from 54.37°S to 43.67°N) including upwelling, shallow lagoon and deep sea stations (**Figure 2** and **Supplementary Table 2**).

Measurements of temperature, conductivity, salinity, depth, pressure, and oxygen were carried out at each station with a vertical profile sampling system (CTD-rosette) and Niskin bottles (Picheral et al., 2014). Chlorophyll *a* concentrations were measured using high-performance liquid chromatography (Van Heukelem and Thomas, 2001; Ras et al., 2008). Phosphate and silicate concentrations were determined using segmented flow analysis (Aminot et al., 2009). Iron concentrations were derived from the biogeochemical model PISCES2 (Aumont et al., 2015). Monthly average estimates of photosynthetically active radiation (PAR) were derived from satellite data[5].

Plotting and Statistical Analysis

All analyses were carried out in R language[6]. Correlation matrices were generated with the *rcorr* function of the *Hmisc* package and plotted using the *corrplot* library. Other graphs were plotted with R library *ggplot2* (Wickham, 2009). Spearman rho correlation analysis were carried out with *cor.test* function.

RESULTS

Diversity and Abundance of Sequences Coding for Diatom Carbon Dioxide Concentration Enzymes

To investigate the diversity and environmental distribution of CCMs in natural populations of diatoms, we searched for sequences coding for CbbX, CA, PEPC, PEPCK, and ME in the eukaryotic unigene catalog of *Tara* Oceans (Carradec et al., 2018) using profile hidden Markov models and sequence similarity networks (see section "Materials and Methods"). The total number of retrieved distinct diatom sequences was: 40 for CbbX, 4,860 for CAs, 943 for PEPC, 488 for PEPCK, and 336 for ME. The obtained CA sequences corresponded to the following classes: 434 alpha, 39 beta, 1,231 delta, 895 gamma, 1,477 iota, 637 theta, and 147 zeta (**Supplementary Table 1**).

We then retrieved the metagenomic and metatranscriptomic read abundances of these sequences across the four main eukaryotic size fractions (**Figure 3A** and **Supplementary Figure 3** and **Supplementary Table 3**). CAs were dominant both in gene number and transcript abundance, with almost one order of magnitude higher levels than the other enzymes under study (**Figure 3A**). CAs comprise on average the 0.2% of the total diatom metatranscriptomic reads, which is similar to the values of all nuclear-encoded subunits of photosystem II

[3]https://doi.org/10.1594/PANGAEA.875582

[4]https://doi.pangaea.de/10.1594/PANGAEA.875567

[5]https://modis.gsfc.nasa.gov/

[6]http://www.r-project.org/

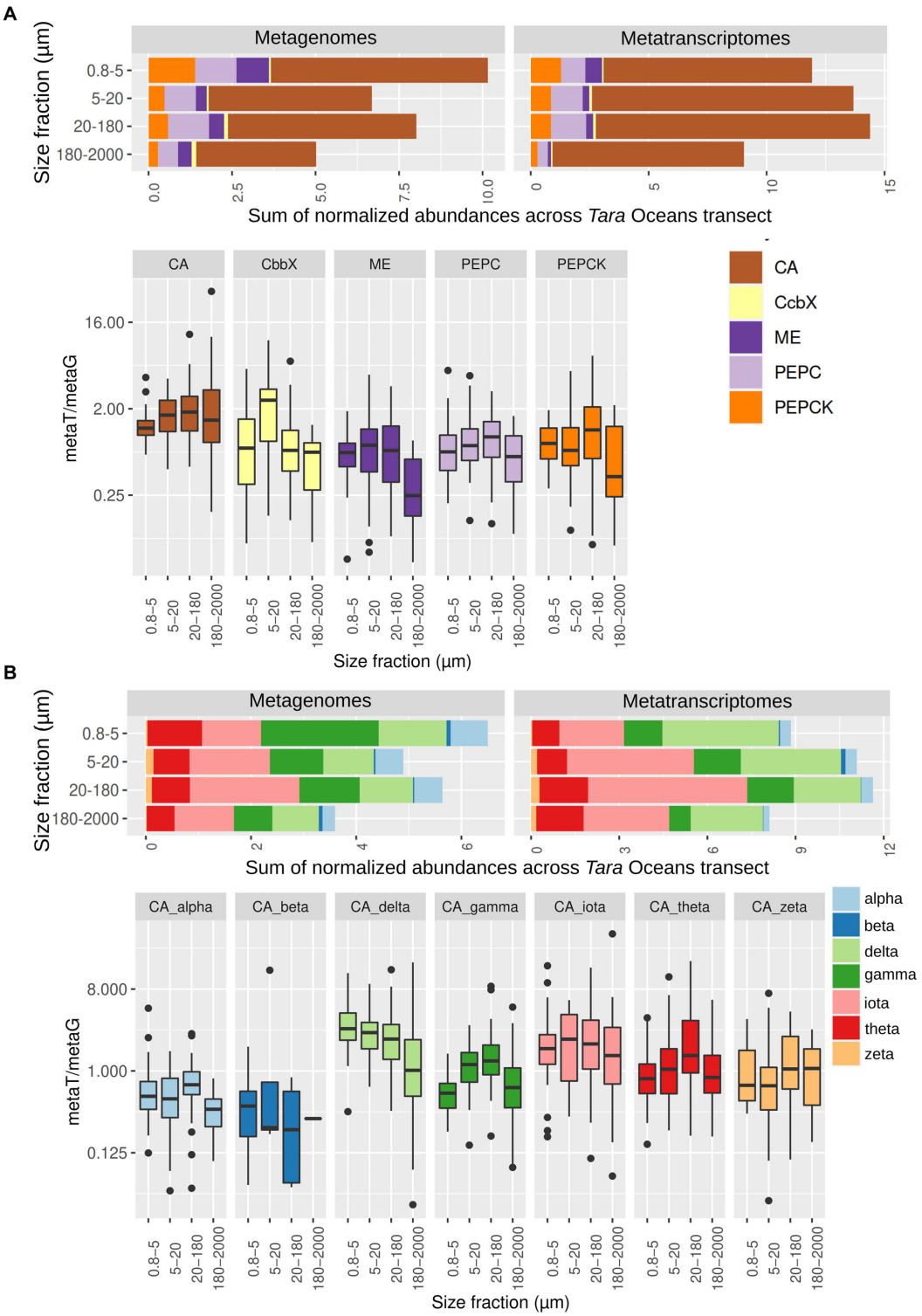


FIGURE 3 | Abundance of genes and transcripts potentially involved in diatom carbon dioxide concentration mechanisms across the different size-fractionated seawater samples collected during the *Tara* Oceans transect. **(A)** Gene and transcript abundances for the five enzymes under study. Barplots show the sum of normalized abundances for all samples in a given size fraction. Boxplots show the gene expression levels based on the abundance ratio in metatranscriptomes and metagenomes (metaT/metaG), and are displayed in logarithmic scale. Abbreviations: carbonic anhydrase (CA), Rubisco activase (CbbX), malic enzyme (ME), phosphoenolpyruvate carboxylase (PEPC), and phosphoenolpyruvate carboxykinase (PEPCK). **(B)** Gene and transcript abundances for the different types of CAs. Barplots and boxplots are displayed as indicated in panel **(A)**.

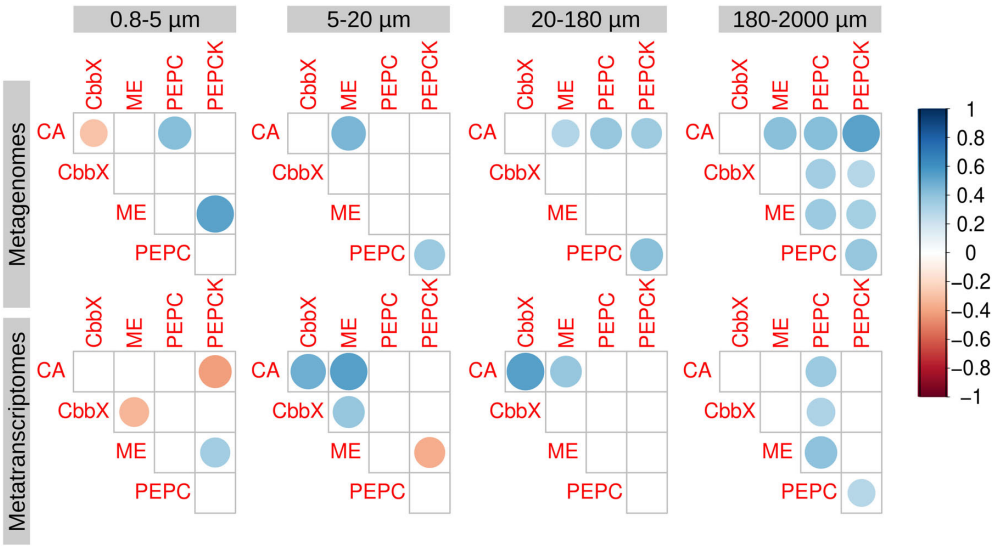


FIGURE 4 | Correlation analysis between the diatom genes and transcripts potentially involved in carbon dioxide concentration mechanisms. Circle size and color intensity are proportional to the Spearman's rho correlation coefficients. Empty spaces refer to non-significant correlation values (two.tailed p-value > 0.05). carbonic anhydrase (CA), Rubisco activase (CbbX), malic enzyme (ME), phosphoenolpyruvate carboxylase (PEPC), and phosphoenolpyruvate carboxykinase (PEPCK).

(**Supplementary Figure 3B**). These results emphasize the importance of CAs in diatom CCMs. For the five enzymes, we found differences between size fractions, probably related with differential needs for maintaining CCMs according to cell sizes and/or aggregation forms: while CbbX gene and transcript abundance increases when moving toward the bigger size fractions, the opposite is observed for the other enzymes (**Figure 3A** and **Supplementary Figure 3**).

Among the different classes of CAs (**Figure 3B** and **Supplementary Figure 4**), delta, gamma and iota are the most abundant (18–37% and 9–47% of the total CA gene and transcript abundance, respectively, with the percentage range corresponding to the minimum and maximum values depending on the size fraction), followed by theta (13–16% and 9–10%) and alpha (7–11% and 2–4%), whereas zeta and beta represent <2% of gene or transcript abundance. Iota-CA showed the highest gene abundances, and the highest transcript abundances together with delta-CA.The CA classes show differences in abundance between metagenomes and metatranscriptomes, reflecting differences in the expression levels of their genes (**Figure 3B**). Delta CA is the most expressed and shows a clear expression increase toward the smaller size classes. It is followed by iota, whose expression does not vary between size fractions. On the opposite, alpha and beta are the least expressed classes.

We also analyzed the correlations between the transcript abundances of the different enzymes (**Figure 4**). In general, we did not find strong correlations in expression of the potential components of a biochemical CCM: ME, PEPC, and PEPCK (**Figure 4**). An exception was nonetheless noted in the largest size fraction (180–2,000 µm) (**Figure 4**), where epizoic and large chain-forming diatoms are found. Thus, this pathway cannot be discarded, but it seems clear that it would not be universal in diatom communities.

Biogeographical Distribution of Genes and Transcripts of Diatom Carbon Dioxide Concentration Enzymes Show Abundance Hotspots

We plotted the biogeographical abundance distributions of the genes and transcripts under study (**Figure 5** and **Supplementary Figures 5**, **6**). All enzymes show a widespread occurrence, but with some regional patterns in abundance. A clear regional pattern is found for PEPCK, which shows its lowest gene and transcript abundances in the Southern Ocean (SO) across all size fractions. In addition, we detected several stations that can be considered abundance hotspots for the genes and/or the transcripts coding for carbon concentrating enzymes, but showing divergence between size fractions, pointing to the effect of cell size and/or aggregation.

For CA, the highest gene and transcript abundances were detected in the Indian and North Atlantic Oceans (IO and NAO, respectively) as well as in a few stations in the South Atlantic Ocean (SAO; **Figure 6** and **Supplementary Figures 7**, **8**). The most abundant CA classes are widespread in the global ocean (but with some differences in their abundances). On the contrary, the low-abundant zeta and beta classes are mainly detected outside the equatorial region (**Figure 6**).

Correlations Between the Environmental Variables and Genes Encoding Diatom Carbon Dioxide Concentration Enzymes

We carried out a correlation analysis between gene and transcript abundances of the enzymes under study and the

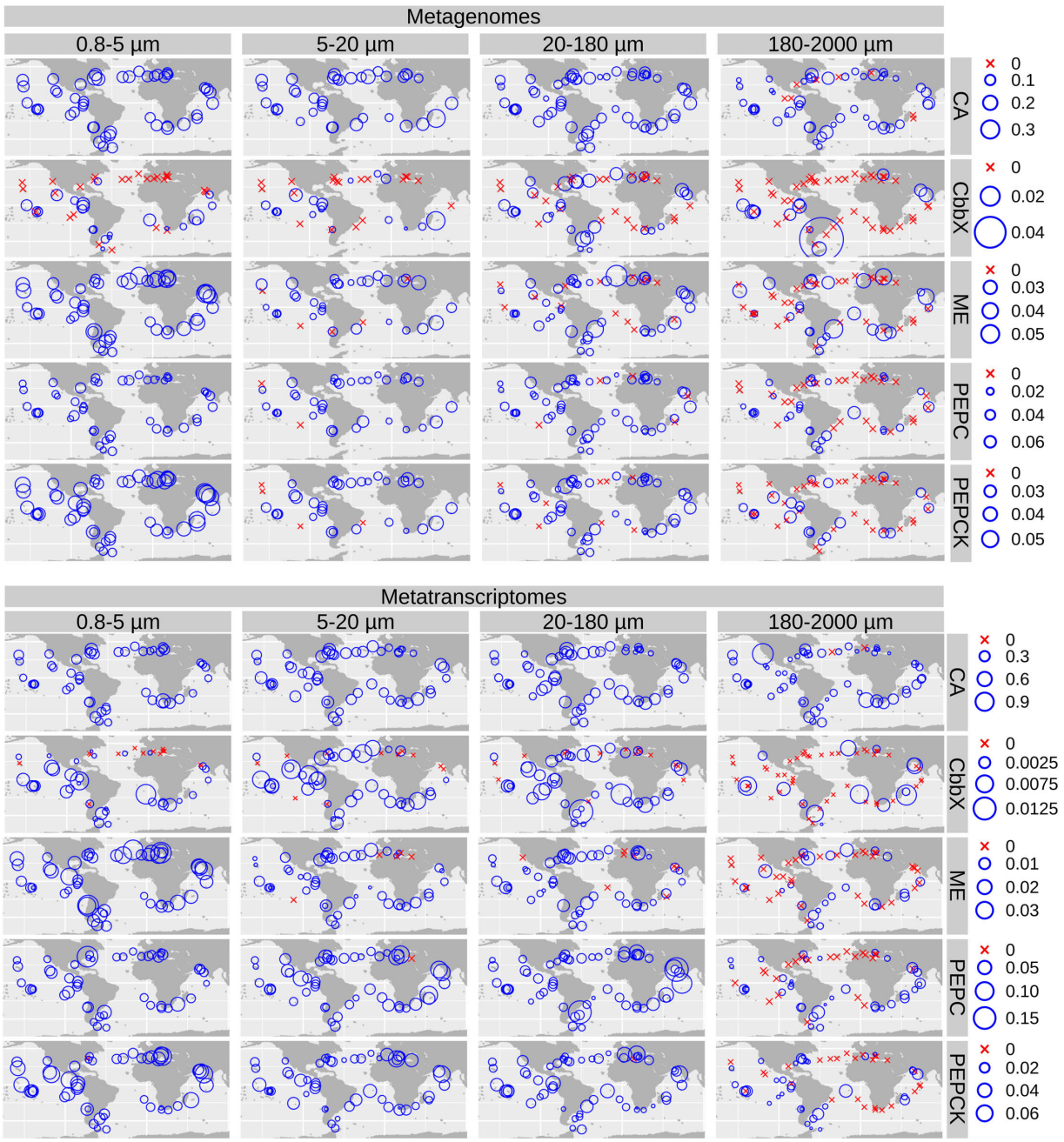


FIGURE 5 | Biogeography of genes and transcripts potentially involved in diatom carbon dioxide concentration mechanisms. Circle sizes are proportional to the gene or transcript abundance (% of total diatom gene or transcript read abundance), while crosses indicate absence of detection. Color code varies according to the size fractions. CA, carbonic anhydrase; CbbX, Rubisco activase; ME, malic enzyme; PEPC, phosphoenolpyruvate carboxylase, PEPCK, phosphoenolpyruvate carboxykinase.

physicochemical and carbon chemistry variables (**Figure 7**). Many of these variables are correlated among each other (**Figure 7A**), which has to be taken into account when interpreting the patterns.

When focusing on transcript abundances, PEPCK and CbbX displayed an anticorrelation with absolute latitude, whereas ME and most of the CA classes showed the opposite (**Figure 7B**). These patterns can be related to the effect of temperature and/or PAR, or the fact that in the current dataset the absolute latitude is linked to nutrient and carbon chemistry variables (**Figure 7A**). CbbX is correlated with phosphate, as are many CA classes.

The correlation matrix with the carbon chemistry variables and CCM enzymes are displayed in **Figure 7C**. The trends

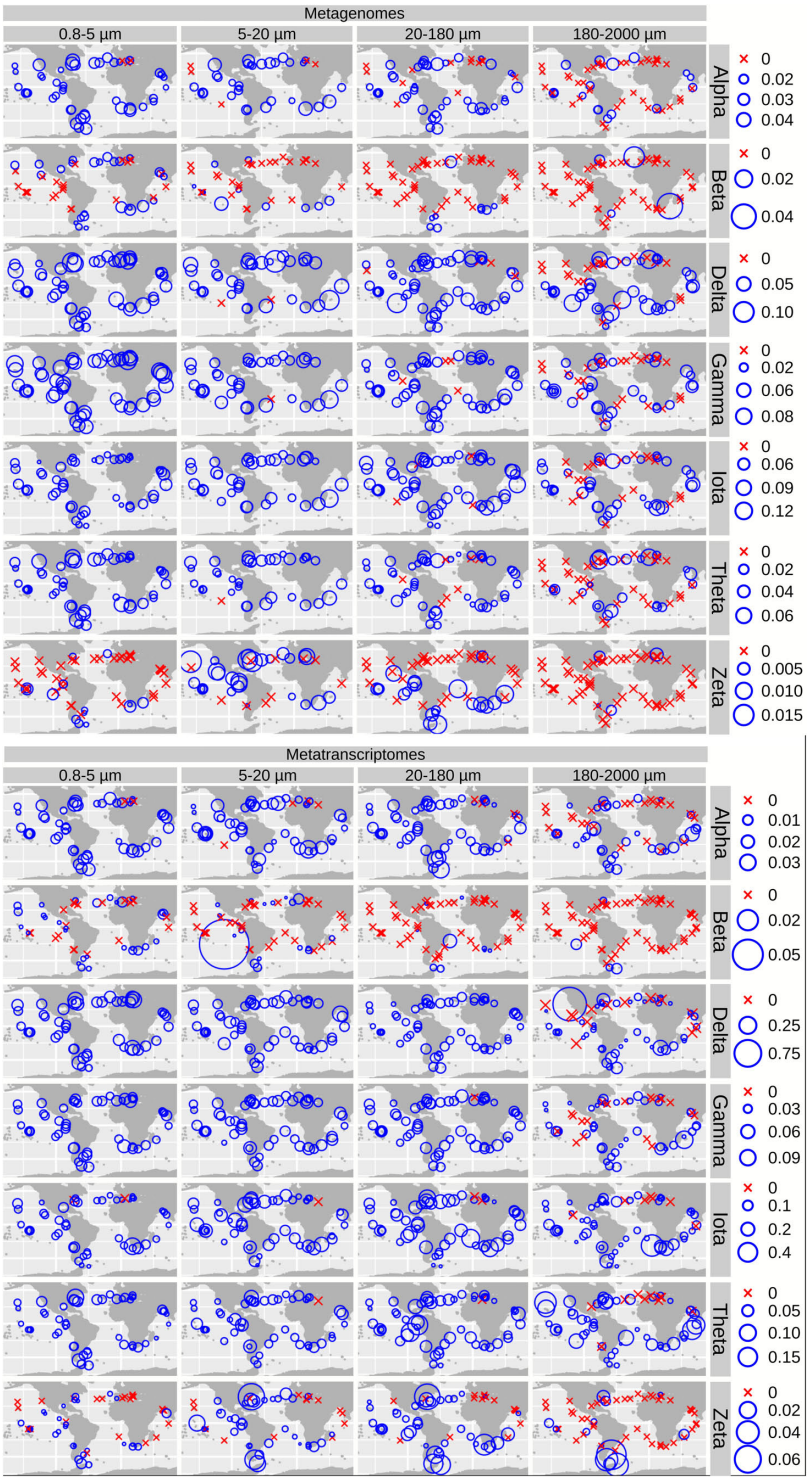


FIGURE 6 | Biogeographical patterns for the diatom genes and transcripts encoding carbonic anhydrases. Circle sizes are proportional to the gene or transcript abundance (% of total diatom gene or transcript abundance), while crosses indicate absence of detection.

revealed that the partial pressure of CO_2 displayed no correlation with PEPC transcript abundance in any size fraction. By contrast, PEPCK showed strong positive correlations in two size fractions (0.8–5 and 20–180 µm) with the partial pressure of CO_2 and strong negative correlations with pH. On the contrary, ME was significantly negatively correlated with CO_2 and positively with

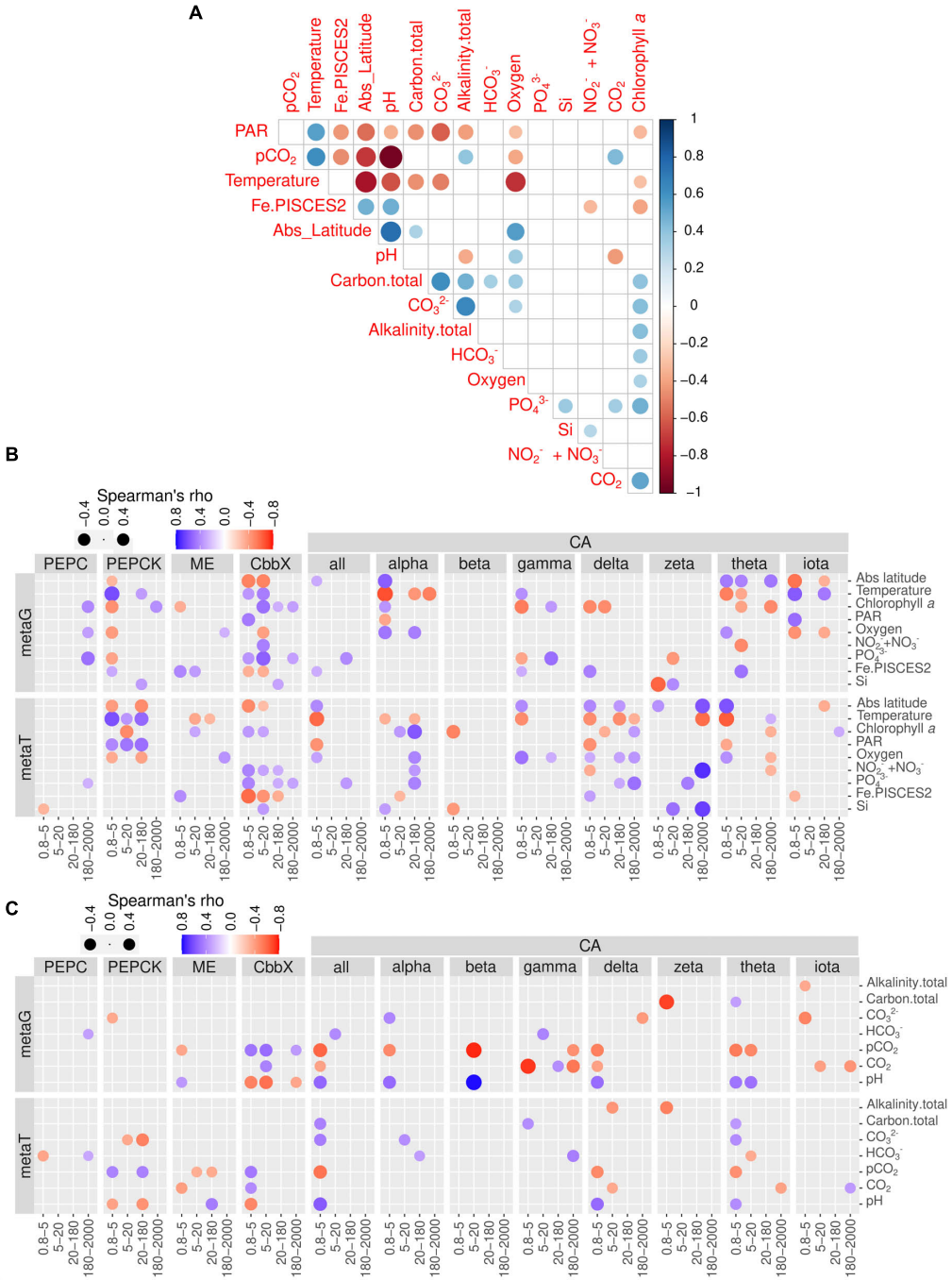


FIGURE 7 | Environmental distribution of genes and transcript potentially involved in diatom carbon dioxide concentration mechanisms. **(A)** Pairwise correlation of the matrix of contextual parameters. **(B)** Correlations of nutrients, chlorophyll *a* and temperature with gene and transcript abundances for the enzymes under study. **(C)** Correlations of carbonate chemistry measurements with gene and transcript abundances for the enzymes under study. Circle color varies according to Spearman rho's correlation coefficient, while size varies according to the absolute value of the coefficient. Only statistically significant correlations are displayed (two-tailed test, $p < 0.05$). PAR, photosynthetically active radiation; CA, carbonic anhydrase; CbbX, Rubisco activase; ME, malic enzyme; PEPC, phosphoenolpyruvate carboxylase, PEPCK, phosphoenolpyruvate carboxykinase.

pH. Interestingly, CbbX, the least expressed enzyme of diatom CCM, showed significant positive correlations with CO_2 (partial pressure and concentrations) and negatively varied with pH only in the smallest size fractions.

CAs in general displayed strong positive correlations with bicarbonate, carbonate ion concentrations, as well as total alkalinity, and negatively correlated with the partial pressure of CO_2 only in the smallest size class. Specifically, delta and theta

classes show strong negative correlations with the partial pressure of CO_2 for smaller size groups. Surprisingly, iota-CA, one of the most abundant CAs, was generally not well correlated with the carbon chemistry variables. Similarly, beta, gamma and zeta-CA did not show any clear trends with carbon chemistry parameters. Zeta-CA gene expression levels for the largest size class exhibited strong positive correlations with absolute latitude, Si and NO_3- + NO_2- levels and varied inversely with temperature (**Figure 7B**). The expression levels of alpha, delta, gama, and theta for the smallest size class were negatively correlated with temperature and hence the average expression level for all CAs also indicated a similar trend.

DISCUSSION

CbbX

The identification of CbbX and its functional role as a Rubisco activation system in diatoms were reported less than a decade ago (Mueller-Cajar et al., 2011) and very little information is available from natural diatom assemblages. We present here the first baseline data regarding the natural variability of this important protein.

The number of CbbX sequences was very low compared with the other sequences. This can be related to the fact that the *Tara* Oceans gene catalog corresponds to assembled sequences from transcriptomes of polyadenylated RNA (Alberti et al., 2017; Carradec et al., 2018), thus minimizing the detection of plastid-encoded versions of CbbX. In addition chloroplast sequences were removed from the final catalog (Carradec et al., 2018), which might also filter the nuclear-encoded versions of CbbX due to its similarity to the plastid encoded versions (Bhat et al., 2017).

The metatranscriptomic read abundance for the sequences coding for CbbX was also very low. *A priori*, a high expression would be expected if we consider the ability of marine diatoms to fix one fifth of global carbon fixation per year and that Rubisco is the most abundant protein on the planet. However, this low total metatranscriptomic read abundance is probably an underestimation due to the low number of retrieved sequences, as the expression of these genes (based on the abundance ratio between metranscriptomes and metagenomes) is similar to those of the other enzymes under study (**Figure 3A**). In addition, low transcript abundances do not necessarily imply a low enzymatic activity. It can be possible that the CbbX function in marine diatoms is controlled by both nuclear and plastid-encoded CbbX versions. Moreover, the gene expression for both CbbX and Rubisco can be linearly varied, and hence a low transcript abundance for CbbX would indicate low transcript abundance for Rubisco. Indeed, diatoms possess an efficient CCM, thus they do not require a high Rubisco concentration: the amount is <6% of the total cellular protein according to both field and culture experiments (Losh et al., 2013), much less than in land plants. All this information may justify the low transcript levels for CbbX in the current work. This must be particularly true in the oligotrophic open ocean where nitrogen can be limiting because Rubisco plays a role as a nitrogen reservoir (Herrig and Falkowski, 1989). Under nitrogen limitation, the nuclear-encoded proteins are synthesized preferentially over those proteins that are encoded in the plastid (Herrig and Falkowski, 1989). In this contest, it is worth mentioning the significant positive correlation between transcript levels for CbbX and NO_3^- + NO_2^- concentrations. The strongest correlation was observed for the smallest size fraction while no correlation was detected for the largest size range, which can be related to the fact that smaller diatoms allocate lesser nitrogen resources to build Rubisco than the large centric diatoms (Wu et al., 2014). Finally, the negative correlation of gene and transcript abundance for CbbX against Fe concentrations may indicate its higher activity in the open ocean iron-limited areas. Under nitrogen limitation, the cellular demand for Fe can be significantly low since Fe is essential for nitrogen metabolism. Small-sized diatoms growing under low nitrogen and Fe limited area can allocate less amount of nitrogen resource to synthesize Rubisco and this could be an evolutionary strategy for the open ocean diatoms.

Our analysis shows that the gene abundance and expression levels of CbbX were positively correlated with pCO_2 and negatively related with pH in the smallest size fraction (0.8–5 μm). This trend suggests that within the smallest diatoms from the global ocean, CbbX activation is likely to be a prominent feature. CbbX homologs have been detected in the model diatoms *T. pseudonana* and *Phaeodactylum tricornutum,* as well as *Asterionella formosa* and other stramenopiles (Jensen et al., 2017). Nevertheless, it has been already shown that the quantitative level of Rubisco protein does not represent the rate of carboxylation (Raines, 2003; Gontero and Salvucci, 2014). This is likely because the rate of carboxylation is controlled principally through the activation of Rubisco by a motor protein like CbbX (Jensen et al., 2017). However, such a conformational change of Rubisco from an inactive protein to its active form involves several factors and is more complicated than was initially presumed (Gontero and Salvucci, 2014). Therefore, our observation reveals the likely significance of CbbX protein in diatoms. Furthermore, the structure and activation mechanism of RCA in higher plants or green algal lineages are considerably different from the CbbX protein found in red algal lineage taxa characterized by ID type Rubisco. Based on such information, it has been hypothesized that these two different types of motor proteins for Rubisco activation probably resulted from convergent evolution coupled with changing atmospheric CO_2/O_2 levels (Mueller-Cajar et al., 2014).

Based on the recently reported high diversity of diatom CCMs (Young et al., 2016; Iñiguez et al., 2020) and the efficiency of the Rubisco 1D type, it has been postulated that diatom CCMs and Rubisco might have co-evolved (Young and Hopkinson, 2017) with changing environmental variables like decreasing CO_2 and increasing O_2 levels (Reinfelder, 2011; Clement et al., 2017). The photorespiratory energy loss is relatively lower in marine diatoms than in other phytoplankton (Rech et al., 2008), while the specificity factor (τ) of Rubisco for CO_2 relative to O_2 is considerably higher (Tortell, 2000). This strengthens the fact that diatoms are capable of maintaining a high CO_2:O_2 ratio in the vicinity of Rubisco through active DIC pumping systems (Reinfelder, 2011). The main evolutionary diversification

in marine diatoms took place during the time when atmospheric CO_2 levels dropped significantly (Reinfelder et al., 2000) and therefore diatoms among the other phytoplankton groups are likely to have developed the most efficient CCMs and Rubisco type (Young et al., 2012). This type ID Rubisco from red algal lineage can perform its highest activity under low CO_2:O_2 ratio and demands low nutrients as well as energy investment in a CCM; this was likely to be the key factor for mass expansion of diatoms and coccolithophores in the Phanerozoic oceans under very high O_2 and low CO_2 levels (Rickaby and Hubbard, 2019).

At the heart of the CCM of diatoms and other algae is the pyrenoid (Badger et al., 1998), a spherical structure in the chloroplast stroma consisting of a matrix of tightly packed Rubisco and RCA. The molecular mechanism by which Rubisco aggregates to form the pyrenoid matrix was recently resolved in the model green alga *Chlamydomonas reinhardtii*, where a low-complexity repeat protein, Essential Pyrenoid Component 1 (EPYC1), links Rubisco to form the pyrenoid (Mackinder et al., 2016). The primary sequences of disordered proteins like EPYC1 are known to evolve rapidly compared with those of structured proteins, but their physicochemical properties are under selective pressure and are evolutionarily maintained. Therefore, Mackinder et al., 2016 searched for proteins with similar physicochemical properties (i.e., repeat number, length, high isoelectric point, disorder profile, and absence of transmembrane domains) across a broad range of algae. They found potential EPYC1-like proteins in the diatoms *T. pseudonana* and *P. tricornutum*, which do not exhibit sequence conservation between them. Expectelly, a BLAST search using these sequences against the MATOU-v1 catalog did not retrieve any similar sequences (data not shown).

Carbonic Anhydrases

Carbonic anhydrases are one of the highest upregulated CCM enzymes in diatom cells grown in CO_2 limited conditions (Clement et al., 2017), however, CAs also play several other physiological roles apart from photosynthesis (Raven, 1995). Out of eight different types of CAs (Jensen et al., 2020), seven subclasses of CAs are found to be constitutively expressed in diatoms (Samukawa et al., 2014; Jensen et al., 2019). The present study also noticed the presence of the expressed genes of all eight types of CAs in the natural diatom populations from the surface ocean. Such high variability and abundance of CAs in diatoms are quite exclusive relative to other organisms and could be due to their evolutionary complexity. The fact that CA transcript levels are the highest in the *Tara* Oceans dataset also explains its profound role in CCMs in marine natural populations of diatoms and indicates that diatoms in the global oceans are likely to be operating a biophysical CCM. Marine diatoms usually show very high intercellular conversion of bicarbonate to CO_2 and vice-versa to maximize CO_2 levels in the vicinity of Rubisco and reduce the diffusive loss of CO_2 from the cell (Matsuda and Kroth, 2014) and hence the significance of CAs are eminently important. Zeng et al. (2019) noticed a strong correlation between Rubisco and CA activities in the model marine diatom *P. tricornutum* and suggests that the rate of carboxylation is directly dependent on the rate of DIC supply which is mediated by CA.

The subcellular location of different CAs can be directly linked to CO_2 acquisition. There are some isoforms which are found in the diatom chloroplast, such as iota-CA, beta-CA and theta-CA (Tanaka et al., 2005; Kikutani et al., 2016). The proximity of such CAs to Rubisco probably results in a more efficient CO_2 acquisition. Consistent with this view, our observation of a significant negative correlation between gene and transcript abundances of theta-CA against pCO_2 for the smallest size fraction also points to an upregulated function of this enzyme at low pCO_2 levels. The presence of the chloroplast-targeted theta-CA in some haptophyte species suggest that the diatom ancestor might have acquired this CA gene via horizontal gene transfer (Nonoyama et al., 2019). Regarding iota-CA, there are many gene copies coding for chloroplast-targeted iota-CAs in common marine diatoms like Odontella, but in a few other species the gene is absent (Nonoyama et al., 2019).

The recent research by Clement et al. (2017) reported the regulation of the latest type of CA, known as "Low CO_2 inducible protein of 63kDa" or LCIP63 in the marine diatom *T. pseudonana*. Later Jensen et al. (2019) confirmed its biochemical function as a CA and renamed it as iota-CA, also showing its widespread occurrence in the *Tara* Oceans dataset. Most importantly, the authors also reported that this type of CA showed its highest expression in surface waters and decreased with increasing depths. It should therefore also be noted that CCMs are likely to play a role in energy dissipation to remove extra energy from the cells and hence, under light inhibition in surface waters, CCMs involving this iota-CA might be used both for carbon acquisition as well as energy dissipation (Kroth et al., 2008). With increasing depth, light stress reduces and CO_2 levels increase and therefore, the need for running a CCM involving iota-CA may be reduced. Our results also found that the highest abundant and expressed gene was iota-CA within marine diatoms from surface waters of the global ocean.

The absence of any significant correlation between iota-CA and carbon chemistry in general probably suggests that this enzyme functions despite CO_2 variability in surface waters. This shows that the expression levels of CAs may not necessarily be coupled with CO_2 levels. For example, in the coccolithophore *Emiliania huxleyi* the transcript of a delta-CA can exhibit high levels of expression irrespective of CO_2 variability (Soto et al., 2006).

Carbonic anhydrase-zeta showed its highest expression at high latitudes for 180–2,000 μm size and seemed to be associated with larger diatoms. The positive correlations with $NO_2^- + NO_3^-$ and Si levels also support this view since the large-celled diatoms in high latitude regions are usually found within eutrophic waters because they have very low surface area to volume ratios.

Within the smallest size fraction (0.8–5 μm) the positive correlation between CA gene expression and pH (coupled with negative correlation with pCO_2) indicates that under high pH smaller diatoms use CA in their CCMs.

Our results also show that CAs are ubiquitous among all size classes of diatoms, and display high diversity. The abundance and expression of different types of CAs can largely be impacted by

trace metal availability in the sea. Importantly, marine diatoms showed the ability to replace a specific metal ion with other more available forms under metal limited conditions (Lane et al., 2005). These metalloenzymes mostly use zinc (Zn) as a cofactor, but other metals such as cadmium (Cd), cobalt (Co), iron (Fe), and manganese (Mn) have also been reported to be associated with different CAs (Morel et al., 2020). In fact the Zn-CAs have been identified to substitute Zn with Co and Cd in surface waters (Morel et al., 2020). In the present study, out of these seven CAs detected, alpha, beta and theta-CAs use Zn ions, whereas, gama, delta and zeta showed the ability to substitute Zn with other metal ions including Cd, Co, and even Fe (Jensen et al., 2020). The highest expression (i.e., metatranscriptomic to metagenomic abundance ratios) were seen in those CAs which are capable of replacing Zn with other metals (**Figure 3B**). The recently identified iota-CA contains Mn and the availability of Mn can be much higher than Zn, particularly in coastal regions. Hence, marine diatoms might have selectively used this particular Mn-containing CA to cope with the available metal ions. However, this will remain a topic for future research to correlate different CA abundance and expression with trace metal concentrations in the global oceans.

The C4 Enzymes

Our analyses revealed that the transcripts for the enzymes of the putative C4 biochemical CCM did not display co-occurring profiles, with the exeption of the largest size fraction (180–2,000 μm). It has to be noted that this size fraction has a prevalence of copepods, considered one of the most abundant multicellular organisms on the planet, and thus the sequencing signal from diatoms is weaker than in the other size fractions. This can be reflected by the higher variability in this size fraction with respect to the absence/presence of diatom genes and transcripts in the different sampling sites. Therefore, we cannot extend so far the speculations about this biochemical pathway, but it seems clear that the process is unlikely to be prevalent in natural communities, as the transcript levels for the three enzymes of a potential biochemical CCM were significantly lower than CA.

There are many experimental studies on marine diatoms showing the expression of all C4 enzymes (Reinfelder et al., 2000, 2004, Reinfelder, 2011; Roberts et al., 2007b), however their active functioning was not confirmed. The negative correlation between gene expression levels of ME and pCO_2/fugacity (as well as the positive correlation with pH) suggests that under CO_2 limitation the diatoms are likely to use this enzyme (except in the largest size fraction). Clement et al. (2017) observed that ME showed the lowest activity among all C4 enzymes and the ratio of Rubisco to PEPC was persistently >1 in the experimental marine diatoms. Our results are also consistent with this observation. Haimovich-Dayan et al. (2013) conducted an experiment by genetically silencing an essential C4 enzyme (pyruvate−orthophosphate dikinase, PPDK) in *P. tricornutum* and observed no major reduction in carboxylation rate. The authors concluded that marine diatoms are likely to use a C4 CCM for dissipating extra light energy. In another study by McGinn and Morel (2008), it was noticed that inhibition of two C4 enzymes (PEPC and PEPCK) resulted in significant reduction in photosynthetic activity in three model marine diatoms. There was almost no study available from any natural diatom population on this aspect and therefore, this study confirms that the relative contribution of C4 CCMs in surface water diatoms is significantly lower than C3 CCM. Moreover, a detailed investigation on deep chlorophyll maxima diatoms is essential to have a clearer picture about functioning of C4 CCM in marine diatoms. Furthermore, additional information on bicarbonate transporter proteins would also shed more light on this topic.

There is some experimental evidence showing higher resilience of phytoplankton communities to increasing CO_2 levels from the oceanic region within the "subtropical north and above" (Schulz et al., 2013; Holding et al., 2015; Hoppe et al., 2017 and references therein). Diatoms from the Arctic and other high latitude seas showed high resilience to variable CO_2 levels (Feng et al., 2009; Hoppe et al., 2018a,b; Sett et al., 2018; Wolf et al., 2018, 2019). This suggests that certain diatom species have high physiological plasticity to tackle the problem of increasing CO_2 levels and therefore, no alteration in photosynthetic performance or growth rate was noticed in relation to changing CO_2 levels in the experimental simulations (Hoppe et al., 2018a; Hoppe et al., 2018b; Wolf et al., 2018; Biswas et al., unpublished data). The diatoms from this region are likely to possess a constitutive CCM and therefore variable CO_2 levels did not reveal any correlation with the gene expression of CbbX protein and other enzymes.

Hoppe et al. (2018a,b) and Wolf et al. (2018), and Biswas et al. (unpublished data) showed that Arctic diatoms are also highly resilient to the combined stress of irradiance and CO_2 levels. This suggests that they have highly evolved cellular mechanisms to counteract photo-inhibition mechanisms. Unpublished data from Biswas et al. showed that an Arctic diatom has high plasticity to control pigment synthesis to combat light limitation/inhibition. Likewise, active functioning of CCM in the surface waters also could be used for these diatoms and the expression levels of C4 enzymes as well as CA can be high. Low latitude phytoplankton may face a stronger impact of photo-inhibition, particularly in the surface waters than the high latitude groups (Tortell, 2000). Hence, the cells living in surface waters may trade off cellular energy between photo-protection and carboxylation. In that case, CA gene expression may be high on the surface. Light is never limiting in this region and hence light dependent DIC uptake can never be hampered. In an experimental study by Biswas et al. (2017) on a tropical coastal diatom community, it was noticed that when light and CO_2 both became limiting, carboxylation significantly hampered and resulted in low organic carbon accumulation. On the other hand, under saturated light the signature of non-photochemical quenching was noticed, even though carbon biomass accumulation was higher. Moreover, there is a continuous need of photosystem repair in the surface water due to the breakdown of the D1 protein of pigment system-II (Lavaud et al., 2016). A CCM, either C3- or C4-like mechanism could also be used for dissipating extra light energy in the surface waters (Haimovich-Dayan et al., 2013). It is also possible that a functional CCM in diatom cells from this region may help alleviate light stress and allow

photosynthetic performance to remain unaffected. The recent study by Jensen et al. (2019) showed that iota-CA showed the highest expression in surface waters and decreased with increasing depth. Light/energy limitation in the subsurface water may be the reason for such down regulation (Kroth et al., 2008).

CONCLUSION

This is the first attempt to assess the diversity, abundance, and distribution of CCMs in natural diatom assemblies at a global ocean scale. We carried out paired metagenomic and metatranscriptomic analyses, targeting five key enzymes, including components of the physical pathway as well as components associated with the putative biochemical mechanism.

We observed changes in transcript abundances in the different size fractions depending on the enzymes, pointing to the effect of different cell sizes and/or aggregation forms, such as chains.

CA was the most abundant and highly expressed gene with almost an order of magnitude higher values than the remaining enzymes, thus confirming the importance of biophysical CCM in natural diatom communities. Among the different classes of this enzyme, the most prevalent was the iota class, which was only recently characterized as a CA (Jensen et al., 2019) and so the information presented here represents the first data on its abundance in natural diatom assemblages.

Biogeographical and environmental distributions showed a complex pattern of responses to CO_2 levels, total phytoplankton biomass, temperature and nutrient concentrations. This is in part due to the current limitations in the dataset, such as the correlations between different environmental variables or the poor representation of certain conditions. The future generation of data from new regions (e.g., Arctic Ocean) can ameliorate these limitations. It is nevertheless expected to obtain complex patterns when assessing the bulk responses of natural diatom populations, since species can differ in their physiological and molecular responses to the environment.

The transcript levels for the three enzymes of a potential biochemical CCM were significantly lower than CA. In addition, we did not find strong correlations among them, except in the largest size fraction (180–2,000 μm), where epizoic and large chain-forming diatoms are found. Thus, while the biochemical pathway cannot be excluded, it seems clear that the process is unlikely to be prevalent in natural communities.

Overall, this work provides a snapshot of diatom CCMs in the global ocean, providing valuable information toward the prediction of diatom responses in an ocean under anthropogenic change.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HB and CB designed the project. JJPK carried out the bioinformatic analysis. JJPK, CB, and HB analyzed the results and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the FFEM—French Facility for Global Environment, French Government 'Investissements d'Avenir' programmes OCEANOMICS (ANR-11-BTBR-0008), FRANCE GENOMIQUE (ANR-10-INBS-09-08), MEMO LIFE (ANR-10-LABX-54), PSL Research University (ANR-11-IDEX-0001-02), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Diatomic; grant agreement No. 835067), Agence Nationale de la Recherche "Phytomet" (ANR-16-CE01-0008), and BrownCut (ANR-19-CE20-0020) projects. JJPK acknowledges postdoctoral funding from the Fonds Français pour l'Environnement Mondial. This article is contribution number 115 of *Tara* Oceans.

ACKNOWLEDGMENTS

We would like to thank all colleagues from the *Tara* Oceans consortium as well as the Tara Ocean Foundation for their inspirational vision. We also acknowledge Dr. Shruti Malviya and Dr. Federico Ibarbalz for their scientific input during the development of this project and Ms. Saumya Silori for helping with the station location map. We are also grateful to the two reviewers for their useful comments. HB is also thankful to the Director NIO, for his kind support to carry out this research. NIO contribution number is 6711.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.657821/full#supplementary-material

Supplementary Figure 1 | Sequence analysis of CbbX and homologs. **(A)** Protein similarity network for the Pfam domain AAA (PF00004). Each node represents a given sequence and those sequences with similarity higher than a score cutoff are linked (score cut-off of 40 in blast alignment). The network was built with sequences retrieved from the literature and from reference genomes and transcriptomes. Nodes are colored according to their taxonomy. The cluster containing CbbX sequences and close homologs is circled. **(B)** Phylogeny of the Pfam domain AAA from the sequences belonging to the cluster highlighted in panel **(A)**. The branch for CbbX is colored in yellow, whereas the remaining back branches are annotated as stage V sporulation protein K. **(C)** Phylogeny of the Pfam domain AAA from CbbX sequences, corresponding to the branch highlighted in panel **(B)**. Color code varies according to the taxonomy. The sequence similarity network and the phylogenies were used as references for the selection of *Tara* Oceans unigenes encoding diatom CcbX. The list of sequences and the alignment are available in **Supplementary Table 1**.

Supplementary Figure 2 | Sequence analysis of iota carbonic anhydrase and homologs. Protein similarity network for the Pfam domain CaMKII_AD (PF08332). Each node represents a given sequence and those sequences with similarity higher than a score cutoff are linked (score cut-off of 18 in blast alignment). The network was built with sequences retrieved from the literature and from reference genomes and transcriptomes, as well as *Tara* Oceans unigenes. Nodes are colored according to their taxonomy. The cluster containing the iota carbonic anhydrase is encircled, as well as the cluster containing calcium/calmodulin-dependent protein kinase II. The list of sequences is available in **Supplementary Table 1**.

Supplementary Figure 3 | Relative abundance of genes and transcripts potentially involved in diatom carbon dioxide concentration mechanisms in comparison to genes involved in other metabolisms. **(A)** Sum of normalized abundances for all samples in a given size fraction. **(B)** Gene and transcript abundances. Values in the box plots correspond to the% of total diatom gene or transcript abundance in the corresponding sample, and are displayed in $\log_2$ scale. In order to compare with other pathways, we also show the abundances for ribosomal proteins and for the nuclear-encoded subunits of photosystem II. Abbreviations: PSII, photosystem II; CA, carbonic anhydrase; CcbX, Rubisco activase; ME, malic enzyme; PEPC, phosphoenolpyruvate carboxylase; PEPCK, phosphoenolpyruvate carboxykinase.

Supplementary Figure 4 | Relative abundance of genes and transcripts coding for the different classes of diatom carbonic anhydrases across the size-fractionated seawater samples collected during the *Tara* Oceans transect. **(A)** Sum of normalized abundances for all samples in a given size fraction. **(B)** Gene and transcript abundances. Values in the box plots correspond to the% of total diatom gene or transcript abundance, and are displayed in $\log_2$ scale.

Supplementary Figure 5 | Biogeographical distribution of genes potentially involved in diatom carbon dioxide concentration mechanisms. Barplots are proportional to the gene abundance (% of the total diatom gene read abundance), while color indicates the enzyme: CA, carbonic anhydrase; CbbX, Rubisco activase; ME, malic enzyme; PEPC, phosphoenolpyruvate carboxylase, PEPCK, phosphoenolpyruvate carboxykinase. The Y axis shows the *Tara* Oceans stations and the ocean regions: MS, Mediterranean Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean.

Supplementary Figure 6 | Biogeographical distribution of transcripts potentially involved in diatom carbon dioxide concentration mechanisms. Barplots are proportional to the transcript abundance (% of the total diatom transcript read abundance), while color indicates the enzyme. Abbreviations: CA, carbonic anhydrase; CbbX, Rubisco activase; ME, malic enzyme; PEPC, phosphoenolpyruvate carboxylase, PEPCK, phosphoenolpyruvate carboxykinase.

Supplementary Figure 7 | Biogeographical distribution of genes coding for the different classes of diatom carbonic anhydrases. Barplots are proportional to the gene abundance (% of the total diatom gene read abundance), while color indicates the carbonic anhydrase class. The Y axis shows the *Tara* Oceans stations and the ocean regions. Abbreviations: MS, Mediterranean Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean.

Supplementary Figure 8 | Biogeographical distribution of transcripts coding for the different classes of diatom carbonic anhydrases. Barplots are proportional to the transcript abundance (% of the total diatom transcript read abundance), while color indicates the carbonic anhydrase class. The Y axis shows the *Tara* Oceans stations and the ocean regions. Abbreviations: MS, Mediterranean Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean.

Supplementary Table 1 | Sequences used in the current work. **(A)** List of Pfam models. **(B)** MATOU-v1 unigenes from diatoms coding for the five analyzed enzymes. **(C)** Sequences used for **Supplementary Figure 1A**. **(D)** Sequences and alignments used for **Supplementary Figure 1B**. **(E)** Sequences and alignments used for **Supplementary Figure 1C**. **(F)** Sequences used for **Supplementary Figure 2**.

Supplementary Table 2 | Contextual data for the *Tara* Oceans samples used in the current work. Original source: https://doi.org/10.1594/PANGAEA.875582 and https://doi.pangaea.de/10.1594/PANGAEA.875567.

Supplementary Table 3 | Abundance of genes and transcripts potentially involved in diatom carbon dioxide concentration mechanisms across the different *Tara* Oceans samples. Values correspond to the% of total diatom read abundance (in rpkm). carbonic anhydrase (CA), Rubisco activase (CbbX), malic enzyme (ME), phosphoenolpyruvate carboxylase (PEPC), and phosphoenolpyruvate carboxykinase (PEPCK).

REFERENCES

Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., et al. (2017). Viral to metazoan marine plankton nucleotide sequences from the *Tara* Oceans expedition. *Sci. Data* 4:170093.

Aminot, A., Kérouel, R., and Coverly, S. C. (2009). "Nutrients in seawater using segmented flow analysis," in *Practical Guidelines for the Analysis of Seawater*, ed. O. Wurl (Boca Raton, FL: CRC Press), 143–178.

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature* 459, 185–192. doi: 10.1038/nature08057

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015). PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. *Geosci. Model Dev.* 8, 2465–2513. doi: 10.5194/gmd-8-2465-2015

Badger, M. (2003). The roles of carbonic anhydrases in photosynthetic CO_2 concentrating mechanisms. *Photosynth. Res.* 77:83.

Badger, M. R., Andrews, T. J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W., et al. (1998). The diversity and coevolution of rubisco, plastids, pyrenoids, and chloroplast-based CO_2-concentrating mechanisms in algae. *Can. J. Bot.* 76, 1052–1071. doi: 10.1139/cjb-76-6-1052

Bar-On, Y. M., and Milo, R. (2019). The global mass and average rate of rubisco. *Proc. Natl. Acad. Sci. U.S.A.* 116, 4738–4743. doi: 10.1073/pnas.1816654116

Bhat, J. Y., Thieulin-Pardo, G., Hartl, F. U., and Hayer-Hartl, M. (2017). Rubisco activases: AAA+ chaperones adapted to enzyme repair. *Front. Mol. Biosci.* 4:20. doi: 10.3389/fmolb.2017.00020

Biswas, H., Shaik, A. U. R., Bandyopadhyay, D., and Chowdhury, N. (2017). CO_2 induced growth response in a diatom dominated phytoplankton community from SW Bay of Bengal coastal water. *Estuar. Coast. Shelf Sci.* 198, 29–42. doi: 10.1016/j.ecss.2017.07.022

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244.

Boyd, P. W., Claustre, H., Levy, M., Siegel, D. A., and Weber, T. (2019). Multi-faceted particle pumps drive carbon sequestration in the ocean. *Nature* 568, 327–335. doi: 10.1038/s41586-019-1098-2

Busseni, G., Rocha Jimenez Vieira, F., Amato, A., Pelletier, E., Pierella Karlusich, J. J., Ferrante, M. I., et al. (2019). Meta-omics reveals genetic flexibility of diatom nitrogen transporters in response to environmental changes. *Mol. Biol. Evol.* 36, 2522–2535. doi: 10.1093/molbev/msz157

Cacefo, V., Ribas, A. F., Zilliani, R. R., Neris, D. M., Domingues, D. S., Moro, A. L., et al. (2019). Decarboxylation mechanisms of C4 photosynthesis in *Saccharum* spp. increased PEPCK activity under water-limiting conditions. *BMC Plant Biol.* 19:144. doi: 10.1186/s12870-019-1745-7

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9:373.

Chen, I. M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2018). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, 666–677.

Clement, R., Dimnet, L., Maberly, S. C., and Gontero, B. (2016). The nature of the CO_2−concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. *New Phytol.* 209, 1417–1427. doi: 10.1111/nph.13728

Clement, R., Lignon, S., Mansuelle, P., Jensen, E., Pophillat, M., Lebrun, R., et al. (2017). Responses of the marine diatom *Thalassiosira pseudonana* to changes in CO_2 concentration: a proteomic approach. *Sci. Rep.* 7:42333.

Davis, A., Abbriano, R., Smith, S. R., and Hildebrand, M. (2017). Clarification of photorespiratory processes and the role of malic enzymes in diatoms. *Protist* 168, 134–153. doi: 10.1016/j.protis.2016.10.005

Dorrell, R. G., Gile, G., Mccallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., et al. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* 6:e23717.

Dorrell, R. G., Villain, A., Perez-Lamarque, B., de Kerdrel, G. A., McCallum, G., Watson, A. K., et al. (2021). Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2009974118. doi: 10.1073/pnas.2009974118

Edmond, J. M. (1970). High precision determination of titration alkalinity and total carbon dioxide content of sea water by potentiometric titration. *Deep-Sea Res. Oceanogr. Abstr.* 17, 737–750. doi: 10.1016/0011-7471(70)90038-0

Endo, H., Sugie, K., Yoshimura, T., and Suzuki, K. (2015). Effects of CO_2 and iron availability on *rbcL* gene expression in Bering sea diatoms. *Biogeosciences* 12, 2247–2259. doi: 10.5194/bg-12-2247-2015

Erb, T. J., and Zarzycki, J. (2018). A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr. Opin. Biotechnol.* 49, 100–107. doi: 10.1016/j.copbio.2017.07.017

Falkowski, P. G., and Raven, J. A. (2013). *Aquatic Photosynthesis*. Princeton, NJ: Princeton University Press.

Feng, Y., Hare, C. E., Leblanc, K., Rose, J. M., Zhang, Y., DiTullio, G. R., et al. (2009). Effects of increased pCO_2 and temperature on the North Atlantic spring bloom. I. The phytoplankton community and biogeochemical response. *Mar. Ecol. Prog. Ser.* 388, 13–25. doi: 10.3354/meps08133

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281, 237–240. doi: 10.1126/science.281.5374.237

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2020. *Earth Syst. Sci. Data* 12, 3269–3340.

Gontero, B., and Salvucci, M. E. (2014). Regulation of photosynthetic carbon metabolism in aquatic and terrestrial organisms by rubisco activase, redox-modulation and CP12. *Aquat. Bot.* 118, 14–23. doi: 10.1016/j.aquabot.2014.05.011

Granum, E., Raven, J. A., and Leegood, R. C. (2005). How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Can. J. Bot.* 83, 898–908. doi: 10.1139/b05-077

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gutknecht, J., Bisson, M. A., and Tosteson, F. C. (1977). Diffusion of carbon dioxide through lipid bilayer membranes. Effects of carbonic anhydrase, bicarbonate, and unstirred layers. *J. Gen. Physiol.* 69:779. doi: 10.1085/jgp.69.6.779

Haimovich-Dayan, M., Garfinkel, N., Ewe, D., Marcus, Y., Gruber, A., Wagner, H., et al. (2013). The role of C4 metabolism in the marine diatom *Phaeodactylum tricornutum*. *New Phytol.* 197, 177–185.

Hennon, G. M., Ashworth, J., Groussman, R. D., Berthiaume, C., Morales, R. L., Baliga, N. S., et al. (2015). Diatom acclimation to elevated CO_2 via cAMP signalling and coordinated gene expression. *Nat. Clim. Chang.* 5:761. doi: 10.1038/nclimate2683

Herrig, R., and Falkowski, P. G. (1989). Nitrogen limitation in *Isochrysis galbana* (haptophyceae). I. Photosynthetic energy conversion and growth efficiencies.I. *J. Phycol.* 25, 462–471. doi: 10.1111/j.1529-8817.1989.tb00251.x

Holding, J. M., Duarte, C. M., Sanz-Martín, M., Mesa, E., Arrieta, J. M., Chierici, M., et al. (2015). Temperature dependence of CO_2-enhanced primary production in the European Arctic Ocean. *Nat. Clim. Chang.* 5, 1079–1082. doi: 10.1038/nclimate2768

Hopkinson, B. M., Dupont, C. L., Allen, A. E., and Morel, F. M. (2011). Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3830–3837. doi: 10.1073/pnas.1018062108

Hopkinson, B. M., Dupont, C. L., and Matsuda, Y. (2016). The physiology and genetics of CO_2 concentrating mechanisms in model diatoms. *Curr. Opin. Plant Biol.* 31, 51–57. doi: 10.1016/j.pbi.2016.03.013

Hoppe, C. J., Schuback, N., Semeniuk, D. M., Maldonado, M. T., and Rost, B. (2017). Functional redundancy facilitates resilience of subarctic phytoplankton assemblages toward ocean acidification and high irradiance. *Front. Mar. Sci.* 4:229. doi: 10.3389/fmars.2017.00229

Hoppe, C. J. M., Schuback, N., Semeniuk, D., Giesbrecht, K., Mol, J., Thomas, H., et al. (2018a). Resistance of arctic phytoplankton to ocean acidification and enhanced irradiance. *Polar Biol.* 41, 399–413. doi: 10.1007/s00300-017-2186-0

Hoppe, C. J. M., Wolf, K. K., Schuback, N., Tortell, P. D., and Rost, B. (2018b). Compensation of ocean acidification effects in arctic phytoplankton assemblages. *Nat. Clim. Change* 8, 529–533. doi: 10.1038/s41558-018-0142-9

Iñiguez, C., Capó−Bauçà, S., Niinemets, Ü, Stoll, H., Aguiló−Nicolau, P., and Galmés, J. (2020). Evolutionary trends in rubisco kinetics and their co−evolution with CO_2 concentrating mechanisms. *Plant J.* 101, 897–918. doi: 10.1111/tpj.14643

Jensen, E., Clément, R., Maberly, S. C., and Gontero, B. (2017). Regulation of the calvin–benson–bassham cycle in the enigmatic diatoms: biochemical and evolutionary variations on an original theme. *Philos. Trans. R. Soc. B* 372:20160401. doi: 10.1098/rstb.2016.0401

Jensen, E. L., Clement, R., Kosta, A., Maberly, S. C., and Gontero, B. (2019). A new widespread subclass of carbonic anhydrase in marine phytoplankton. *ISME J.* 13, 2094–2106. doi: 10.1038/s41396-019-0426-8

Jensen, E. L., Maberly, S. C., and Gontero, B. (2020). Insights on the functions and ecophysiological relevance of the diverse carbonic anhydrases in microalgae. *Int. J. Mol. Sci.* 21:2922. doi: 10.3390/ijms21082922

Jin, X., Gruber, N., Dunne, J. P., Sarmiento, J. L., and Armstrong, R. A. (2006). Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Global Biogeochem. Cycles* 20:GB2015.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Katoh, K., and Toh, H. (2008). Improved accuracy of multiple ncRNA alignment by incorporating structural information into a MAFFT-based framework. *BMC Bioinformatics* 9:212. doi: 10.1186/1471-2105-9-212

Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., 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 Biol.* 12:e1001889. doi: 10.1371/journal.pbio.1001889

Kikutani, S., Nakajima, K., Nagasato, C., Tsuji, Y., Miyatake, A., and Matsuda, Y. (2016). Thylakoid luminal θ-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9828–9833. doi: 10.1073/pnas.1603112113

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jezequel, V., Mock, T., Parker, M. S., et al. (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 3:e1426. doi: 10.1371/journal.pone.0001426

Kustka, A. B., Milligan, A. J., Zheng, H., New, A. M., Gates, C., Bidle, K. D., et al. (2014). Low CO_2 results in a rearrangement of carbon metabolism to support C4 photosynthetic carbon assimilation in *Thalassiosira pseudonana*. *New Phytol.* 204, 507–520. doi: 10.1111/nph.12926

Lane, T. W., Saito, M. A., George, N. G., Pickering, I. J., Prince, R. C., and Morel, F. M. M. (2005). A cadmium enzyme from a marine diatom. *Nature* 435, 42–42. doi: 10.1038/435042a

Lavaud, J., Six, C., and Campbell, D. A. (2016). Photosystem II repair in marine diatoms with contrasting photophysiologies. *Photosynth. Res.* 127, 189–199. doi: 10.1007/s11120-015-0172-3x

Li, G., Brown, C. M., Jeans, J. A., Donaher, N. A., McCarthy, A., and Campbell, D. A. (2015). The nitrogen costs of photosynthesis in a diatom under current and future pCO_2. *New Phytol.* 205, 533–543. doi: 10.1111/nph.13037

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

Loganathan, N., Tsai, Y. C. C., and Mueller-Cajar, O. (2016). Characterization of the heterooligomeric red-type rubisco activase from red algae. *Proc. Natl. Acad. Sci. U.S.A.* 113, 14019–14024. doi: 10.1073/pnas.1610758113

Losh, J. L., Young, J. N., and Morel, F. M. (2013). Rubisco is a small fraction of total protein in marine phytoplankton. *New Phytol.* 198, 52–58. doi: 10.1111/nph.12143

Mackinder, L. C., Meyer, M. T., Mettler-Altmann, T., Chen, V. K., Mitchell, M. C., Caspari, O., et al. (2016). A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5958–5963. doi: 10.1073/pnas.1522866113

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., et al. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1516–E1525.

Matsuda, Y., Hopkinson, B. M., Nakajima, K., Dupont, C. L., and Tsuji, Y. (2017). Mechanisms of carbon dioxide acquisition and CO_2 sensing in marine diatoms: a gateway to carbon metabolism. *Philos. Trans. R. Soc. Lond B.* 372:20160403. doi: 10.1098/rstb.2016.0403

Matsuda, Y., and Kroth, P. G. (2014). "Carbon fixation in diatoms," in *The Structural Basis of Biological Energy Generation*, ed. M. Hohmann-Marriott (Dordrecht: Springer), 335–362. doi: 10.1007/978-94-017-8742-0_18

McGinn, P. J., and Morel, F. M. (2008). Expression and inhibition of the carboxylating and decarboxylating enzymes in the photosynthetic C4 pathway of marine diatoms. *Plant Physiol.* 146, 300–309. doi: 10.1104/pp.107.110569

Morel, F. M., Lam, P. J., and Saito, M. A. (2020). Trace metal substitution in marine phytoplankton. *Annu. Rev. Earth Planet. Sci.* 48, 491–517. doi: 10.1146/annurev-earth-053018-060108

Morel, F. M. M., Reinfelder, J. R., Roberts, S. B., Chamberlain, C. P., Lee, J. G., and Yee, D. (1994). Zinc and carbon co-limitation of marine phytoplankton. *Nature* 369, 740–742. doi: 10.1038/369740a0

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K., and Bhattacharya, D. (2009). Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* 324, 1724–1726. doi: 10.1126/science.1172983

Mueller-Cajar, O., Stotz, M., and Bracher, A. (2014). Maintaining photosynthetic CO_2 fixation via protein remodelling: the Rubisco activases. *Photosynth. Res.* 119, 191–201. doi: 10.1007/s11120-013-9819-0

Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 479:194. doi: 10.1038/nature10568

Nisumaa, A. M., Pesant, S., Bellerby, R. G. J., Delille, B., Middelburg, J. J., Orr, J. C., et al. (2010). EPOCA/EUR-OCEANS data compilation on the biological and biogeochemical responses to ocean acidification. *Earth Syst. Sci. Data* 2, 167–175. doi: 10.5194/essd-2-167-2010

Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., et al. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules* 9:322. doi: 10.3390/biom9080322

Ohno, N., Inoue, T., Yamashiki, R., Nakajima, K., Kitahara, Y., Ishibashi, M., et al. (2012). CO2-cAMP-responsive cis-elements targeted by a transcription factor with CREB/ATF-like basic zipper domain in the marine diatom Phaeodactylum tricornutum. *Plant Physiol.* 158, 499–513. doi: 10.1104/pp.111.190249

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of *Tara* Oceans data. *Sci. Data* 2:150023.

Picheral, M., Searson, S., Taillandier, V., Bricaud, A., Boss, E., Ras, J., et al. (2014). Vertical profiles of environmental parameters measured on discrete water samples collected with Niskin bottles during the *Tara* Oceans expedition 2009-2013. *PANGAEA* doi: 10.1594/PANGAEA.836319

Pierella Karlusich, J. J., Ibarbalz, F. M., and Bowler, C. (2020). Phytoplankton in the *Tara* Ocean. *Ann. Rev. Mar. Sci.* 12, 233–265. doi: 10.1146/annurev-marine-010419-010706

Pollock, S. V., Colombo, S. L., Prout, D. L., Godfrey, A. C., and Moroney, J. V. (2003). Rubisco activase is required for optimal photosynthesis in the green alga *Chlamydomonas reinhardtii* in a low-CO_2 atmosphere. *Plant Physiol.* 133, 1854–1861. doi: 10.1104/pp.103.032078

Poudel, S., Pike, D. H., Raanan, H., Mancini, J. A., Nanda, V., Rickaby, R. E. M., et al. (2020). Biophysical analysis of the structural evolution of substrate specificity in RuBisCO. *Proc. Natl. Acad. Sci. U.S.A.* 117, 30451–30457. doi: 10.1073/pnas.2018939117

Raines, C. A. (2003). The Calvin cycle revisited. *Photosynth. Res.* 75, 1–10. doi: 10.1163/9789004244672_002

Ras, J., Claustre, H., and Uitz, J. (2008). Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data. *Biogeosciences* 5, 353–369. doi: 10.5194/bg-5-353-2008

Raven, A. (1995). Photosynthetic and non-photosynthetic roles of carbonic anhydrase in algae and cyanobacteria. *Phycologia* 34, 93–101. doi: 10.2216/i0031-8884-34-2-93.1

Rech, M., Morant-Manceau, A., and Tremblin, G. (2008). Carbon fixation and carbonic anhydrase activity in Haslea ostrearia (Bacillariophyceae) in relation to growth irradiance. *Phtosynthetica* 46, 56–62. doi: 10.1007/s11099-008-0011-2

Reinfelder, J. R. (2011). Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Ann. Rev. Mar. Sci.* 3, 291–315. doi: 10.1146/annurev-marine-120709-142720

Reinfelder, J. R., Kraepiel, A. M., and Morel, F. M. (2000). Unicellular C4 photosynthesis in a marine diatom. *Nature* 407, 996–999. doi: 10.1038/35039612

Reinfelder, J. R., Milligan, A. J., and Morel, F. M. (2004). The role of the C4 pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiol.* 135, 2106–2111. doi: 10.1104/pp.104.041319

Rickaby, R. E., and Hubbard, M. E. (2019). Upper ocean oxygenation, evolution of RuBisCO and the Phanerozoic succession of phytoplankton. *Free Radic. Biol. Med.* 140, 295–304. doi: 10.1016/j.freeradbiomed.2019.05.006

Roberts, K., Granum, E., Leegood, R., and Raven, J. A. (2007a). Carbon acquisition by diatoms. *Photosynth. Res.* 93, 79–88. doi: 10.1007/s11120-007-9172-2

Roberts, K., Granum, E., Leegood, R. C., and Raven, J. A. (2007b). C3 and C4 pathways of photosynthetic carbon assimilation in marine diatoms are under genetic, not environmental, control. *Plant Physiol.* 145, 230–235. doi: 10.1104/pp.107.102616

Saade, A., and Bowler, C. (2009). Molecular tools for discovering the secrets of diatoms. *BioScience* 59, 757–765. doi: 10.1525/bio.2009.59.9.7

Samukawa, M., Shen, C., Hopkinson, B. M., and Matsuda, Y. (2014). Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynth. Res.* 121, 235–249. doi: 10.1007/s11120-014-9967-x

Schoefs, B., Hu, H., and Kroth, P. G. (2017). The peculiar carbon metabolism in diatoms. *Phil. Trans. R. Soc. B* 372:20160405. doi: 10.1098/rstb.2016.0405

Schulz, K. G., Bellerby, R. G. J., Brussaard, C. P., Büdenbender, J., Czerny, J., Engel, A., et al. (2013). Temporal biomass dynamics of an Arctic plankton bloom in response to increasing levels of atmospheric carbon dioxide. *Biogeosciences* 10, 161–180. doi: 10.5194/bg-10-161-2013

Sett, S., Schulz, K. G., Bach, L. T., and Riebesell, U. (2018). Shift towards larger diatoms in a natural phytoplankton assemblage under combined high-CO_2 and warming conditions. *J. Plankton Res.* 40, 391–406. doi: 10.1093/plankt/fby018

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Shivhare, D., and Mueller-Cajar, O. (2017). "Rubisco Activase: the molecular chiropractor of the world's most abundant protein," in *Photosynthesis And Bioenergetics*, eds J. Barber and A. V. Ruban (Singapore: World Scientific), 159–187. doi: 10.1142/9789813230309_0008

Smetacek, V. (1999). Diatoms and the ocean carbon cycle. *Protist* 150, 25–32. doi: 10.1016/s1434-4610(99)70006-4

Soto, A. R., Zheng, H., Shoemaker, D., Rodriguez, J., Read, B. A., and Wahlund, T. M. (2006). Identification and preliminary characterization of two cDNAs encoding unique carbonic anhydrases from the marine alga *Emiliania huxleyi*. *Appl. Environ. Microbiol.* 72, 5500–5511. doi: 10.1128/AEM.00237-06

Tanaka, R., Kikutani, S., Mahardika, A., and Matsuda, Y. (2014). Localization of enzymes relating to C4 organic acid metabolisms in the marine diatom, *Thalassiosira pseudonana*. *Photosynth. Res.* 121, 251–263. doi: 10.1007/s11120-014-9968-9

Tanaka, Y., Nakatsuma, D., Harada, H., Ishida, M., and Matsuda, Y. (2005). Localization of soluble beta-carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiol.* 138, 207–217. doi: 10.1104/pp.104.058982

Tortell, P. D. (2000). Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol. Oceanogr.* 45, 744–750. doi: 10.4319/lo.2000.45.3.0744

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., et al. (2018). Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci.* 11, 27–37. doi: 10.1038/s41561-017-0028-x

Van Heukelem, L., and Thomas, C. S. (2001). Computer-assisted high-performance liquid chromatography method development with applications to the isolation and analysis of phytoplankton pigments. *J. Chromatogr. A*. 910, 31–49. doi: 10.1016/s0378-4347(00)00603-4

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer Science & Business Media.

Wilhelm, C., Büchel, C., Fisahn, J., Goss, R., Jakob, T., LaRoche, J., et al. (2006). The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157, 91–124. doi: 10.1016/j.protis.2006.02.003

Wolf, K. K., Hoppe, C. J., and Rost, B. (2018). Resilience by diversity: large intraspecific differences in climate change responses of an Arctic diatom. *Limnol. Oceanogr.* 63, 397–411. doi: 10.1002/lno.10639

Wolf, K. K., Romanelli, E., Rost, B., John, U., Collins, S., Weigand, H., et al. (2019). Company matters: the presence of other genotypes alters traits and intraspecific selection in an Arctic diatom under climate change. *Glob. Chang. Biol.* 25, 2869–2884. doi: 10.1111/gcb.14675

Wu, Y., Jeans, J., Suggett, D. J., Finkel, Z. V., and Campbell, D. A. (2014). Large centric diatoms allocate more cellular nitrogen to photosynthesis to counter slower Rubisco turnover rates. *Front. Mar. Sci.* 1:68. doi: 10.3389/fmars.2014.00068

Young, J. N., Heureux, A. M., Sharwood, R. E., Rickaby, R. E., Morel, F. M., and Whitney, S. M. (2016). Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms. *J. Exp. Bot.* 67, 3445–3456. doi: 10.1093/jxb/erw163

Young, J. N., and Hopkinson, B. M. (2017). The potential for co-evolution of CO_2-concentrating mechanisms and Rubisco in diatoms. *J. Exp. Bot.* 68, 3751–3762. doi: 10.1093/jxb/erx130

Young, J. N., Rickaby, R. E. M., Kapralov, M. V., and Filatov, D. A. (2012). Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. *Philos. Trans. R. Soc. Lond. B* 367, 483–492. doi: 10.1098/rstb.2011.0145

Zallot, R., Oberg, N., and Gerlt, J. A. (2019). The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* 58, 4169–4182. doi: 10.1021/acs.biochem.9b00735

Zeng, X., Jin, P., Xia, J., and Liu, Y. (2019). Correlation of carbonic anhydrase and Rubisco in the growth and photosynthesis in the diatom *Phaeodactylum tricornutum*. *J. Appl. Phycol.* 31, 123–129. doi: 10.1007/s10811-018-1537-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer YM declared a past co-authorship with one of the authors CB to the handling editor.

Copyright © 2021 Pierella Karlusich, Bowler and Biswas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 November 2021
doi: 10.3389/fpls.2021.756421

Metabolite Quantification by Fourier Transform Infrared Spectroscopy in Diatoms: Proof of Concept on *Phaeodactylum tricornutum*

*Matteo Scarsini[1], Adrien Thurotte[1,2], Brigitte Veidl[1], Frederic Amiard[3], Frederick Niepceron[3], Myriam Badawi[1], Fabienne Lagarde[3], Benoît Schoefs[1] and Justine Marchand[1]**

[1] Mer Molécules Santé, Le Mans University, IUML-FR 3473 CNRS, Le Mans, France, [2] Institute of Molecular Biosciences, Goethe University Frankfurt, Frankfurt, Germany, [3] UMR CNRS 6283 Institut des Molécules et des Matériaux du Mans, Le Mans University, Le Mans, France

OPEN ACCESS

Edited by:
Peter J. Lammers,
Arizona State University, United States

Reviewed by:
Alexei E. Solovchenko,
Lomonosov Moscow State University, Russia
Ondrej Prasil,
Institute of Microbiology, Academy of Sciences of the Czech Republic (ASCR), Czechia

***Correspondence:**
Justine Marchand
justine.marchand@univ-lemans.fr

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: 10 August 2021
Accepted: 07 September 2021
Published: 10 November 2021

Citation:
Scarsini M, Thurotte A, Veidl B, Amiard F, Niepceron F, Badawi M, Lagarde F, Schoefs B and Marchand J (2021) Metabolite Quantification by Fourier Transform Infrared Spectroscopy in Diatoms: Proof of Concept on Phaeodactylum tricornutum. Front. Plant Sci. 12:756421. doi: 10.3389/fpls.2021.756421

Diatoms are feedstock for the production of sustainable biocommodities, including biofuel. The biochemical characterization of newly isolated or genetically modified strains is seminal to identify the strains that display interesting features for both research and industrial applications. Biochemical quantification of organic macromolecules cellular quotas are time-consuming methodologies which often require large amount of biological sample. Vibrational spectroscopy is an essential tool applied in several fields of research. A Fourier transform infrared (FTIR) microscopy-based imaging protocol was developed for the simultaneous cellular quota quantification of lipids, carbohydrates, and proteins of the diatom *Phaeodactylum tricornutum*. The low amount of sample required for the quantification allows the high throughput quantification on small volume cultures. A proof of concept was performed (1) on nitrogen-starved experimental cultures and (2) on three different *P. tricornutum* wild-type strains. The results are supported by the observation *in situ* of lipid droplets by confocal and brightfield microscopy. The results show that major differences exist in the regulation of lipid metabolism between ecotypes of *P. tricornutum*.

Keywords: FTIR spectroscopy, *Phaeodactylum tricornutum* Bohlin, macromolecules quantification, algal physiology, lipid droplets, phytoplankton

INTRODUCTION

The increase of human populations, the reduction of farmland and the expansion of cities contribute to an accelerated depletion of natural resources to a point that finding alternative sources of natural resources, including biomolecules, is becoming crucial. Microalgae are a promising feedstock for the production of biomolecules belonging to the major organic macromolecules classes (Mimouni et al., 2012) while utilizing sun energy through the photosynthetic process (Scarsini et al., 2019). In addition, microalga culture does not require farmland, reducing the competition for space with other farm productions (Gordon et al., 2019).

Unfortunately, the current knowledge on microalgae does not allow the development of biotechnological sustainable processes from the ecological and financial point of views. Several bottlenecks need to be addressed before establishing microalgae as efficient cell factories (Vinayak et al., 2015). Obtaining appropriate strains dedicated to a specific production is one of the major challenges. To reach this goal, two strategies may be applied: (1) the exploration of microalgal biodiversity to find appropriate strains with high industrial application potential and (2) the genetic modifications of wild-type microalgal strains to improve their productivity. The efficient and fast selection of natural and genetically improved strains is thus seminal to identify the strains that perform better when cultured. The development of large-scale screening techniques is a challenge. In particular, the identification of cell lines, which possess the phenotype of interest among a mutagenized population, can be a particularly tedious process for which it may be necessary to analyze hundreds or even thousands of transformants.

High throughput methods for microalgal cultivation in small volumes have been developed since a long time (Heinemann and Paulsen, 1999). More recently, Van Wagenen et al. (2014) described a methodology that allows the cultivation of microalgae in small volume multi-well plates with results comparable to a photobioreactor cultivation. However, the development of a macromolecule quantification screening method which does not require large sampling volumes is fundamental since conventional biochemical analyses require relatively high amounts of culture sample.

Fourier transform infrared spectroscopy (FTIR) is a widespread vibrational spectroscopy technique that can provide a chemical fingerprint of samples (either solid, liquid, or gas) with a wide range of applications. In the field of microalgae, FTIR was used to confirm the metabolic plasticity that microalgae display in response to environmental stress such as nutrient stress (Giordano et al., 2001; Heraud et al., 2005; Stehfest et al., 2005; Sigee et al., 2007; Dean et al., 2008; Dean et al., 2010; Wagner et al., 2010; Bajhaiya et al., 2015; Driver et al., 2015; Giordano et al., 2017; Pogorzelec et al., 2017; Cointet et al., 2019). However, most of these studies focused on freshwater microalgae while only few FTIR spectroscopy analyses have been performed on marine microalgae, including diatoms (Giordano et al., 2017; Pogorzelec et al., 2017; Cointet et al., 2019). Among diatoms, *Phaeodactylum tricornutum* is a popular widely used model species, for both physiology studies and genetic engineering (Heydarizadeh et al., 2014, 2017, 2019; Bowler and Falciatore, 2019; Butler et al., 2020). Ten different *P. tricornutum* ecotypes have been isolated around the world, differing on both genotypic and phenotypic levels (Martino et al., 2007; Rastogi et al., 2020). The genetic modification tools available for this species make it possible to create mutants relatively easily (*e.g.*, Nymark et al., 2016; Serif et al., 2018; Stukenberg et al., 2018; Hu and Pan, 2020).

This manuscript reports a high-throughput phenotypic microalgal screening methodology based on FTIR-coupled microscopy to quantify lipids, proteins and carbohydrates cell contents in different growth conditions and different ecotypes or strains of the diatom *P. tricornutum*.

MATERIALS AND METHODS

Microalgal Strains and Culture Conditions

Two ecotypes were used in this study: Pt4 (UTEX 646) and Pt1 (CCMP 2561). Two Pt1 strains were studied, coming from geographically different laboratories: Pt1-Wuhan, kindly provided by Pr H. Hu from the Center for Algal Biology and Applied Research from the Chinese Academy of Sciences at Wuhan (China), and Pt1-Paris kindly provided by Dr. A. Falciatore from the Laboratoire de Biologie du Chloroplaste et Perception de la Lumière chez les Microalgues, Institut de Biologie Physico-Chimique (IBPC, Paris, France).

Both *P. tricornutum* ecotypes were cultured in f/2 medium (Guillard, 1975) supplemented with bicarbonate (14.3 mM $NaHCO_3$). Precultures and experimental cultures were constantly maintained at 21 ± 1°C and 150 μmol m^{-2} s^{-1}. Culture axenicity was checked by plating a drop of culture onto LB medium and f/2 medium supplemented with 0.5% bacto-peptone and 0.1% yeast extract, both incubated in the dark at 21°C for at least 7 days to check for eventual bacterial presence. *P. tricornutum* was cultivated in cell culture flasks (T-25 and T-75 according to the experiment requirement, Sarstedt). To obtain *P. tricornutum* cells (Pt4) enriched in lipids, microalgae were cultivated in nitrogen (N)-depleted f/2 medium (*i.e.* [NO_3^-]=0). To validate the FTIR methodology, a N-starvation experiment was conducted on Pt4 in a 1 L laboratory scale photobioreactor (PBR) (FM150, Photon System Instruments, Czechia) operated in turbidostat mode (∼3.5 × 10^6 cells ml^{-1}) as described in Scarsini (2021). Briefly, the culture was continuously supplied with 1.2 ml min^{-1} of a mixture air/CO_2 (2,000 ppm CO_2) at 21°C and 150 μmol photons s^{-1} m^{-2} (white LED panel). The culture was initially exposed in non-limited conditions (5 days) and then fed with f/2 medium containing 0.15 mM $NaNO_3$ for 20 days. Samplings along the starvation were performed.

A N-starvation experiment (NO_3^- depleted f/2 medium) was finally conducted on Pt1-Wuhan, Pt1-Paris and Pt4: exponential cultures were transferred from N-repleted to N-depleted medium for 18 days in flasks in the same conditions than described above (temperature, light). Samplings were performed along the starvation at 0, 6, 10, and 18 days.

Cell Counting and Growth

Cell density was determined using a Neubauer hemacytometer (Hausser Scientific, Horsham, PA, United States). For the comparison between ecotypes, cell densities were determined using a double beam spectrophotometer (Lambda 25, Perkin Elmer). Optical density (OD) was measured at 735 nm and calibration curves prepared for each strain and each condition in order to associate OD to cell density. Calibration curves are reported in **Supplementary Figure 1A**. Determination of growth rates was obtained using the software CurveExpert Basic[1] through a logistic model fitting (Bolzmann).

[1]https://www.curveexpert.net/

Fourier Transform Infrared Spectroscopy

Samples were centrifugated at 3,000×*g* for 10 min, washed twice with a NaCl solution (1,032 mOsmol), vacuum dried for 5 min and stored at −20°C until use. Samples were resuspended in ultrapure water in order to obtain the desired number of cells ml^{-1}. One µl of the suspension was deposited on a silicon plate and dried at 37°C for 15 min. The silicon plate was previously carefully washed with isopropanol (Fisher Scientific, >99.5%) and dried. Three biological samples (three drops) were considered. Infrared (IR) spectra were collected at the Vibrational Spectroscopy platform of the IMMM (Le Mans University) using a FTIR-microscope Spotlight 200i (Perkin Elmer) controlled by the Spectrum™ software (Perkin Elmer). Spectra were collected from eight randomly distributed spots on the drop avoiding the centre (**Supplementary Figure 2**) with a 4 cm^{-1} spectral resolution and 32 scans per spectrum in transmission modus. Each spectrum was collected in the range 800–4,000 cm^{-1} (**Supplementary Figure 3A**) on a square surface area of 10^4 μm^2. A cut-down spectrum in the range 800–2,000 cm^{-1} (**Supplementary Figure 3B**) was also used. For construction of the baseline, the spectrum was divided into n ranges determined from a convex hull of the spectrum. The minimum Y value of each range was determined. Connecting the minima with spline lines creates the baseline. Starting from "below," a rubber band was stretched over this curve (**Supplementary Figure 3C**). The baseline points that do not lie on the rubber band were discarded. The baseline was subtracted from the raw data to yield the flattened spectrum (**Supplementary Figure 3D**; Shen et al., 2018). To normalize spectra, the Standard Normal Variate (SNV) method was used (Barnes et al., 1989). The SNV is performed first calculating the average intensity value and subsequently subtracting this value from each spectrum (**Supplementary Figure 3E**). Then, the sum of the squared intensities was calculated, and the spectrum divided by the square root of this sum (the standard deviation). After SNV, each spectrum has a mean of 0 and a standard deviation of 1 (offset correction) (**Supplementary Figure 3F**). HyperSpec version 0.99[2], prospectr version 0.2.1[3] (Barnes et al., 1989) and base version 4.0.4 packages were used to perform rubber-band baseline correction, normalization and offset correction, respectively. All packages were compiled with R version 4.1.0. Analyses were performed using the arithmetic mean spectra resulting from three biological replicates (three drops per sample) themselves resulting from eight technical replicates (eight spectra per drop).

Metabolite Quantifications

Sample were centrifugated at 3,000×*g* for 10 min, washed twice with a NaCl solution (1,032 mOsmol) and stored at −20°C until use. Neutral lipid cell quota (called Q_{Lipids}) was quantified by Nile Red staining according to Huang et al. (2019). Room temperature fluorescence spectra were recorded with the Perkin Elmer LS-55 (PerkinElmer®, Excitation wavelength: 530 nm, Emission: 545–800 nm, slit: 5 nm). Dimethyl sulfoxide-diluted glyceryl trioleate (Sigma-Aldrich T7140-500MG, purity ≥ 99%) was used for calibration (**Supplementary Figure 4A**). Soluble proteins and total carbohydrates cell quota (denoted as $Q_{Proteins}$ and $Q_{Carbohydrates}$ respectively) were measured according to Heydarizadeh et al. (2019). Colorimetric analyses were performed using the Lambda-25 spectrophotometer (PerkinElmer®) at 595 and 485 nm (slit: 5 nm) for proteins and carbohydrates, respectively. Bovine serum albumin (Sigma-Aldrich, purity ≥98%) and glucose were used for proteins and carbohydrates quantification calibrations respectively (**Supplementary Figures 4B,C**). Photosynthetic pigments were extracted using acetone (Merck, ≥99.5%) according to Heydarizadeh et al. (2017). Absorbance spectra were recorded between 400 and 800 nm (Perkin Elmer Lambda-25 spectrophotometer) and Chl *a*, Chl c and total carotenoids concentrations calculated according to equations from Heydarizadeh et al. (2017).

Brightfield Microscopy for the Estimation of Cell Size

Pictures were collected with an Olympus CX23 brightfield microscope (objective Olympus 40X) equipped with a ISH300 digital camera (Tucsen Photonics, China). Fiji software[4] (ImageJ v1.53c) was used to measure cell lengths and widths of the different *P. tricornutum* ecotypes and strains.

Confocal Microscopy for the Estimation of Lipid Droplet Volumes

Samples of Pt1 strains and Pt4 were used for imaging using laser scanning confocal microscopy (LSCM). Lipid droplets (LDs) were stained using BODIPY 505/515 (Invitrogen Molecular Probes, Inc., CA, United States), a lipophilic dye with improved properties for confocal imaging. In comparison to Nile Red, the BODIPY 505/515 has a narrower emission spectrum allowing fluorescence enhancement of LDs (Cooper et al., 2010; Govender et al., 2012). One ml of culture (5×10^6 cells ml^{-1}) was treated with 50 µl of DMSO and stained during 10 min with 1 µl of BODIPY 5 mM. The number of cells analyzed for each replicate ranged from 50 to 180 cells (three biological replicates were considered). Three-dimensional projections of LD were performed capturing stack pictures of the sample (400 nm Z resolution) at the Confocal Microscopy platform of the IMMM (Le Mans University) using a LSM800 laser scanning confocal microscope (Carl Zeiss Microscopy GmbH, Germany). All images were acquired using a 63× objective (LCI Plan-Neofluar with numerical aperture of 1.3 water immersion objective DIC M27) with 0.5 as scan zoom. Z sectional images were captured using the 488 and 561 nm line excitation lasers (at respectively 4.5 and 5% laser intensities) for BODIPY and Chl *a* autofluorescence respectively. The BODIPY emission was detected between 410 and 546 nm and Chl *a* autofluorescence between 650 and 700 nm. Z-slice step size was 0.33 µm $slice^{-1}$ increments. Image dimensions were 2,297 × 2,297 pixels with sampling speed of 0.38 µs $pixel^{-1}$ and size of 0.073 µm $pixel^{-1}$. Photomultiplier tubes gain were

[2]https://github.com/r-hyperspec/hyperSpec

[3]https://rdrr.io/cran/prospectr/

[4]https://imagej.net/

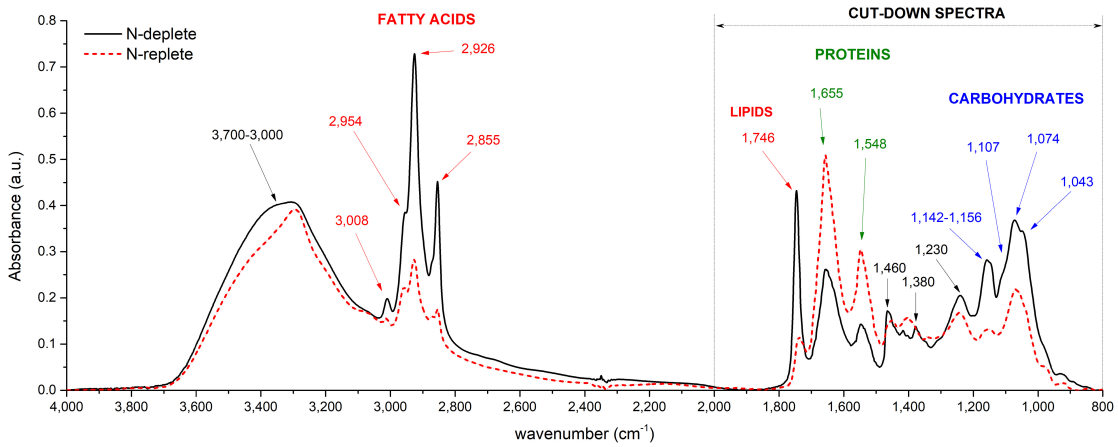


FIGURE 1 | Typical FTIR spectra recorded from *P. tricornutum* (Pt4) in two different conditions: N-replete (blue line) and N-deplete (red line). Major band assignments are indicated by arrows and defined in **Table 1**.

TABLE 1 | Band assignment and functional groups of a typical spectrum of *P. tricornutum*.

Wavenumber (cm^{-1})	Assignment
3,700–3,000	$_{V}$O-H of water, $_{V}$N-H of amide, $_{V}$C-O of carbohydrates
3,008	$_{V}$C-H of C=CH– chains of lipids
2,954	$_{Vas}CH_3$ of methyl groups
2,926	$_{Vas}CH_2$ of methylene groups
2,855	$_{V}CH_2$ and $_{V}CH_3$ of methyl and methylene groups
1,746	$_{V}$C=O ester of lipids and fatty acids
1,655	$_{V}$C=O of proteins (Amide I)
1,548	$_{\delta}$N-H and $_{V}$C–N of proteins (Amide II)
1,460	$_{\delta as}CH_2$ and $_{\delta as}CH_3$ of methyl and methylene groups
1,380	$_{\delta}CH_2$ and $_{\delta}CH_3$ from proteins and $_{\delta}$C–O from carboxylic groups
1,230	$_{Vas}$P=O from phosphodiester of nucleic acids and phospholipids
1,200–900	$_{V}$C-0-C from polysaccharides
980	P-O-P of Polyphosphates

Band assignments are taken from references Giordano et al. (2001) and Mayers et al. (2013). $_{V}$, symmetric stretching; $_{Vas}$, asymmetrical stretching; $_{\delta}$, symmetric deformation (bend); $_{\delta as}$, asymmetric deformation (bend).

manually adjusted to provide optimal brightness and resolution. Fluorescent images were merged and colored using the Zen Blue software (Zeiss). Fiji software[4] (ImageJ v1.53c) was used to estimate the volume of LDs. Stacks from LSCM images containing only the fluorescence channel were converted to a Z-projection. The diameter of each LD was measured manually and used to calculate the volume of a sphere (assuming that all LDs are spherical).

Statistics

Linear and non-linear fitting analyses were performed with the Curve Expert software[1]. Analyses of variance (ANOVA) were performed using a python script exploiting scipy (version 1.5.2), statsmodels (version 0.12.0), and bioinfokit (version 2.0.3).

RESULTS AND DISCUSSION

A high-throughput sensitive methodology based on FTIR was designed and evaluated on the marine diatom *P. tricornutum*. It allowed to assess carbon allocation changes in response to N-starvation.

Typical Infrared Spectra of *Phaeodactylum tricornutum*

The functional groups attached to the different organic macromolecules absorb IR radiation at specific wavelengths, generating a complex spectrum considered as a fingerprint of the cellular composition (Movasaghi et al., 2008; Wagner et al., 2010). For a suitable exploitation, three to four steps are necessary: (1) the FTIR full raw spectrum (800–4,000 cm^{-1}) needs to be processed (**Supplementary Figure 3A**); (2) a cut-down of the spectrum can be further performed keeping only wavenumbers included in 800–2,000 cm^{-1} to avoid the spectral perturbation generated by the large water peak (**Supplementary Figure 3B**); (3) a baseline correction either on the entire spectrum or the cut-down spectrum (**Supplementary Figures 3C,D**); and (4) a normalization procedure can be performed (**Supplementary Figures 3E,F**). FTIR spectra often exhibit baseline offset and curvilinear trend caused by changes in optical path length and light scattering. Baseline removal aims at resetting all spectra on a common baseline (Sandak et al., 2016; Shen et al., 2018). The normalization procedure is carried out in order to uniform the information from spectra and to correct spectra for changes in optical path length and light scattering. This process can be particularly important when dealing with inhomogeneous samples which is the case in microalgae cultures (Sandak et al., 2016). SNV normalization was used to transform each measured spectrum into a signal with a mean of 0 and a variance of 1 (Barnes et al., 1989): the mean value of the spectrum (the offset) (**Supplementary Figure 3E**) is subtracted to each absorbance value, which is subsequently divided by its respective raw standard deviation (**Supplementary Figure 4F**). The data

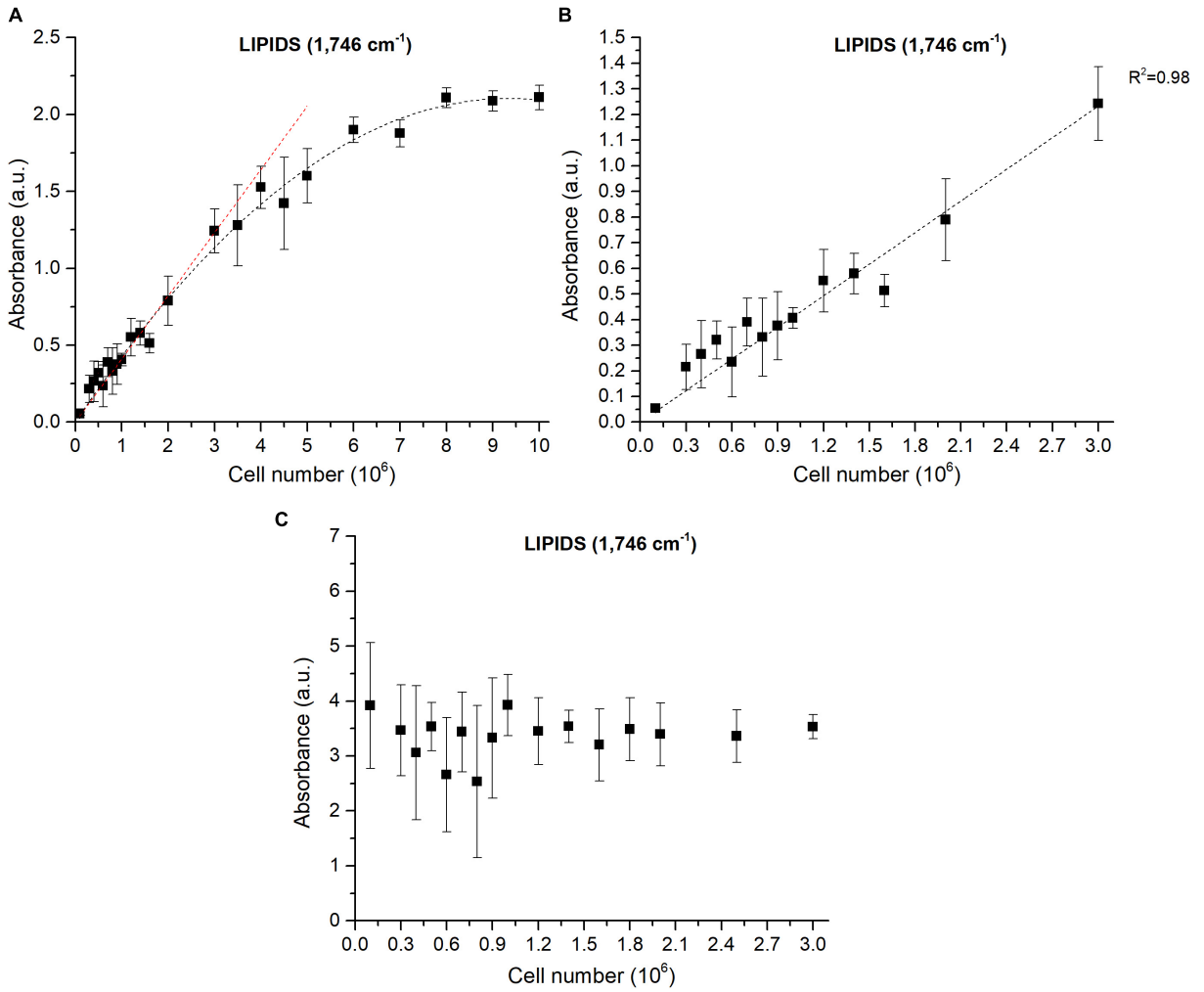


FIGURE 2 | Correlations between FTIR absorbance measured at 1,746 cm^{-1} and the number of cells deposited on the silica plate **(A)** Baseline-corrected spectra considering 0 to 10 million cells, **(B)** Baseline-corrected spectra considering 0 to 3 million cells, **(C)** Baseline-corrected and normalized spectra considering 0 to 3 million cells. The coefficient of determination (R^2) of the linear regression curve is indicated.

obtained in this study are presented in two different ways: (1) baseline-corrected only and (2) baseline-corrected followed by a SNV normalization.

Examples of baseline-corrected spectra obtained from a culture sample of *P. tricornutum* in two different N availability conditions are reported in **Figure 1**. Different peak intensities at particular wavenumbers are observed depending on the N availability in which cells have been grown (**Figure 1**). Two spectral ranges are particularly affected: (1) 800 to 2,000 cm^{-1} and (2) 2,800 to 3,100 cm^{-1}. Strong absorption bands at approximately 1,043, 1,074, 1,107, 1,142, and 1,156 cm^{-1} observed mostly in N-depleted condition correspond to the vibration of C–O bonds of polysaccharides (**Table 1**; Giordano et al., 2001; Mayers et al., 2013). Absorption bands observed at 1,548 and 1,655 cm^{-1} correspond to the vibration of the N-H, C-N and C=O of amide groups belonging to proteins (**Table 1**). The stretching C=O bonds of esters of lipids and fatty acids are found at 1,746 cm^{-1} while those at 2,855, 2,926, and 2,954 cm^{-1} correspond to the symmetric and asymmetric stretching of CH_2 and CH_3 methyl and methylene groups of fatty acids (**Table 1**). As demonstrated by Wagner et al. (2010), absorbance peak height can be used for macromolecules quantification instead of peak integral. On an overall analysis, the N-starvation induces an increase in lipid and carbohydrate cell quotas as well as a decrease in protein cell quota in comparison to N-replete conditions (**Figure 1**), as observed in literature through conventional biochemical analyses (Huang et al., 2019).

Reproducibility of Fourier Transform Infrared Measurements

Reproducibility, as a major principle of a scientific method, is a key feature of an effective analysis. In previous works (Dean et al., 2010; Wagner et al., 2010; Driver et al., 2015), FTIR was performed on a unique sensor and samples required time to dry, making this technique low throughput. In our work, we used imaging spectroscopy that allows to visualize the sample on which the analysis is performed. Multiple areas can thus be selected to record the IR spectra. This allowed to randomly collect absorbance spectra in different areas of the sample *i.e.* recording

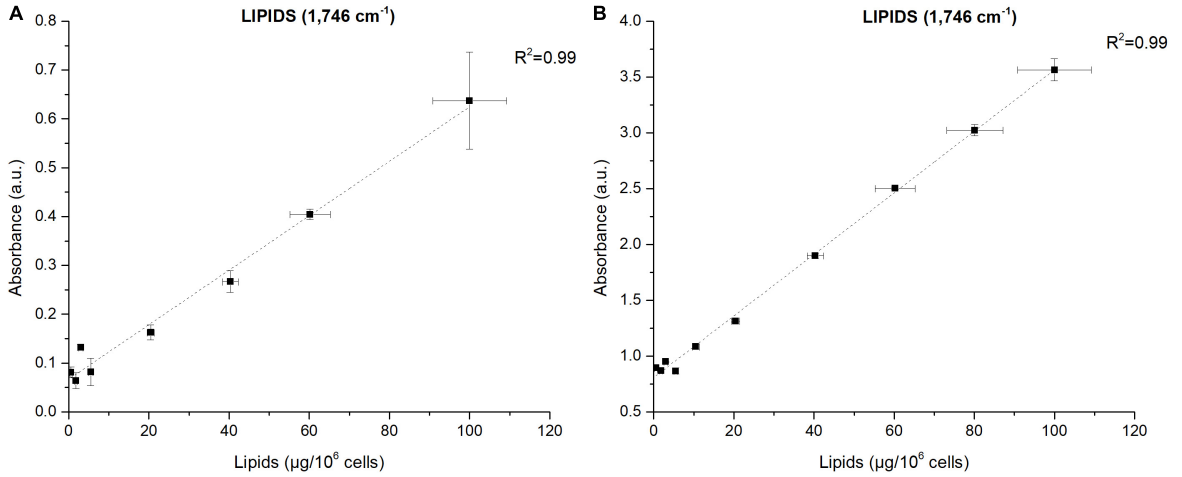


FIGURE 3 | Correlations between FTIR absorbance measured at 1,746 cm^{-1} and lipids concentrations quantified by fluorescence after Nile-Red staining considering **(A)** baseline-corrected spectra and **(B)** baseline-corrected and normalized spectra. Two million cells were deposited on the silica plate. Coefficients of determination (R^2) of the linear regression curves are indicated.

TABLE 2 | Regression equations and coefficients of determination (R^2) of the correlations between FTIR absorbance measured at the different wavenumbers and standard biochemical quantifications considering either baseline-corrected spectra or baseline and normalized-corrected spectra performed on Pt4.

		Baseline spectra		Baseline and Normalized spectra	
Assignments	**Wavenumber (cm^{-1})**	**Regression equations**	***R*²**	**Regression equations**	***R*²**
Carbohydrates	1,043	y=0.054x – 0.012	0.90	y=0.054x – 0.012	0.97
	1,074	y=0.057x + 0.007	0.87	y=0.057x + 0.007	0.97
	1,107	y=0.043x – 0.02	0.89	y=0.043x – 0.02	0.98
	1,142	y=0.044x – 0.04	0.90	y=0.044x – 0.04	0.98
	1,156	y=0.048x – 0.05	0.91	y=0.048x – 0.05	0.98
Proteins	1,548	y=0.061x + 0.08	0.78	y=0.061x + 0.08	0.96
	1,655	y=0.085x + 0.2	0.68	y=0.085x + 0.2	0.96
Lipids	1,746	y=0.0056x + 0.07	0.99	y=0.0056x + 0.07	0.99
	2,855	y=0.0045x + 0.18	0.91	y=0.0045x + 0.18	0.97
	2,926	y=0.0071x + 0.29	0.93	y=0.0071x + 0.29	0.98
	2,954	y=0.0034x + 0.24	0.81	y=0.0034x + 0.24	0.93

y: FTIR absorbance measured at the specific wavenumbers (a.u.). x: concentrations of the different metabolites measured (carbohydrates, proteins, or lipids) with standard biochemical methods.

technical replicates within the same sample (**Supplementary Figure 2**), and different samples within the same holder.

The reproducibility among both technical and biological replicates was tested. Reproducibility of the results increases with the number of technical replicates (**Supplementary Figure 5A**). Indeed, no significant difference (ANOVA, $p<0.05$) was found between the 8 technical replicates for all the wavenumbers studied. However, a higher variability was found (particularly visible for the 1,746 cm^{-1} peak representing lipids) when less than 4 replicates were considered (**Supplementary Figure 5A**). The lack of uniformity of the dried drop on the measuring plate can explain the variability of the measurements (**Supplementary Figure 2**). Eight randomly distributed spots of analyses were set as the optimal compromise between statistical significance, reproducibility and time spent on each sample for the analyses.

Biological replicates consist in samples containing different cells cultivated in the same culture conditions. Six biological replicates (drops A–F; **Supplementary Figures 5B,C**) were tested in order to evaluate the inter-sample reproducibility. The peak amplitude for each biological replicate is the average of the eight technical replicates. From these analyses, different results were observed according to the different spectra processing steps performed. The single baseline correction (**Supplementary Figure 5B**) does not allow a complete reproducibility among the different biological replicates: ANOVA tests ($p<0.05$) revealed differences between peak heights among the biological replicates for all the wavenumbers considered ($p<0.05$). Performing SNV normalization on baseline-corrected spectra (**Supplementary Figure 5C**) reduces the differences among biological replicates and increases the reproducibility of the experiment (no significant difference among replicates, $p<0.05$). Clearly, standardization of spectra using SNV achieves a scaling effect. This normalization was originally proposed to reduce scattering effects in the spectra but was also proved to

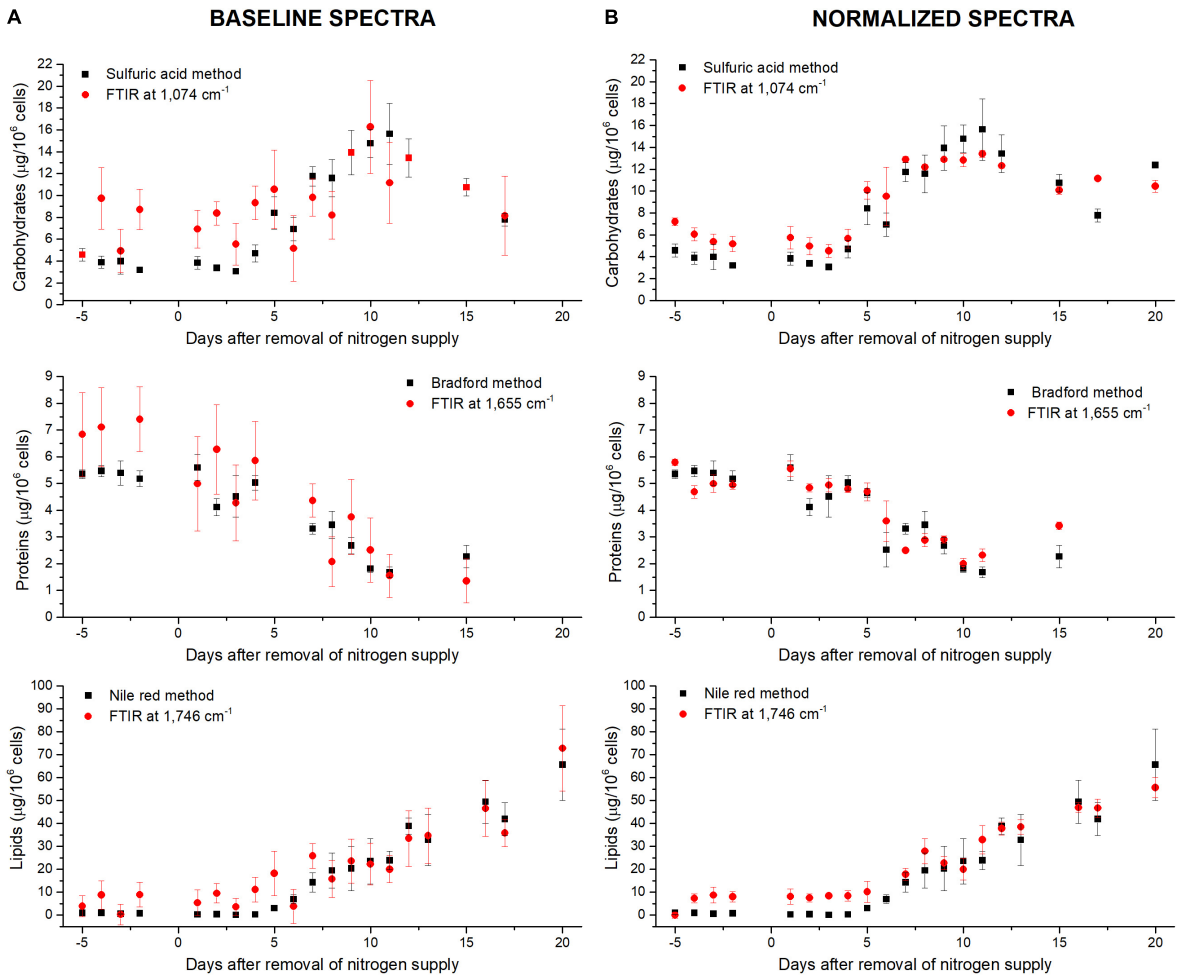


FIGURE 4 | Comparison between standard biochemical and FTIR quantifications along the N-starvation performed in a photobioreactor (Scarsini, 2021) considering **(A)** baseline-corrected spectra or **(B)** baseline and normalized-corrected spectra. FTIR quantifications are performed at 1,074 cm^{-1} for carbohydrates, 1,655 cm^{-1} for proteins and 1,746 cm^{-1} for lipids.

be effective in correcting the interference caused by variations in pellet thickness or/and optical path (Razak et al., 2020).

How Does the Number of Cells Impact the Fourier Transform Infrared Spectroscopy Absorbance?

As already reported in literature, sample thickness is a crucial point and can lead to deviations from the Beer-Lambert's law (Wagner et al., 2010). Cell density is influencing the sample thickness and must thus be taken into account. For a given cell density, the size of the deposited drop can be considered constant since it depends on the force of gravity, the surface tension coefficient and the roughness of the surface of the silica plate (Eid et al., 2018), which are constant in the experiments described in this paper. The diameter of the deposited drop was not impacted during the evaporation of the water as reported previously by Mampallil (2014). To verify the effect of cell density on the experiment reproducibility, FTIR analyses were performed on different cellular densities of Pt4 cells enriched in lipids. Different cell densities were prepared and a drop of each solution deposited on the silica reading plate. As expected, the absorbance increased with cell density and reached a saturation level for densities higher than 6×10^6 cells ml^{-1} (**Figure 2A** and **Supplementary Figure 6A**). The correlation between absorbance and number of cells remained linear under 3×10^6 cells for all the wavenumbers considered (**Figure 2B** and **Supplementary Figure 6B**). **Figure 2C** has been obtained through the SNV normalization procedure to remove the inhomogeneity of the sample (which in this case is the cell density). The graph shows how the replicability at 1,746 cm^{-1} with a number of cells deposited ranging from 10^6 to 3×10^6 cells is high. This is also the case for the other wavenumbers, although less pronounced (**Supplementary Figure 6C**). The instability at lower cellular densities is probably linked to the inhomogeneous distribution of cells inside the drop area (**Supplementary Figure 7**). The stable values observed with the normalization procedure indicates that depositing a specific number of cells is not necessary when remaining between 10^6 and 3×10^6 cells. The normalization step corrects sampling and human errors when cell density is

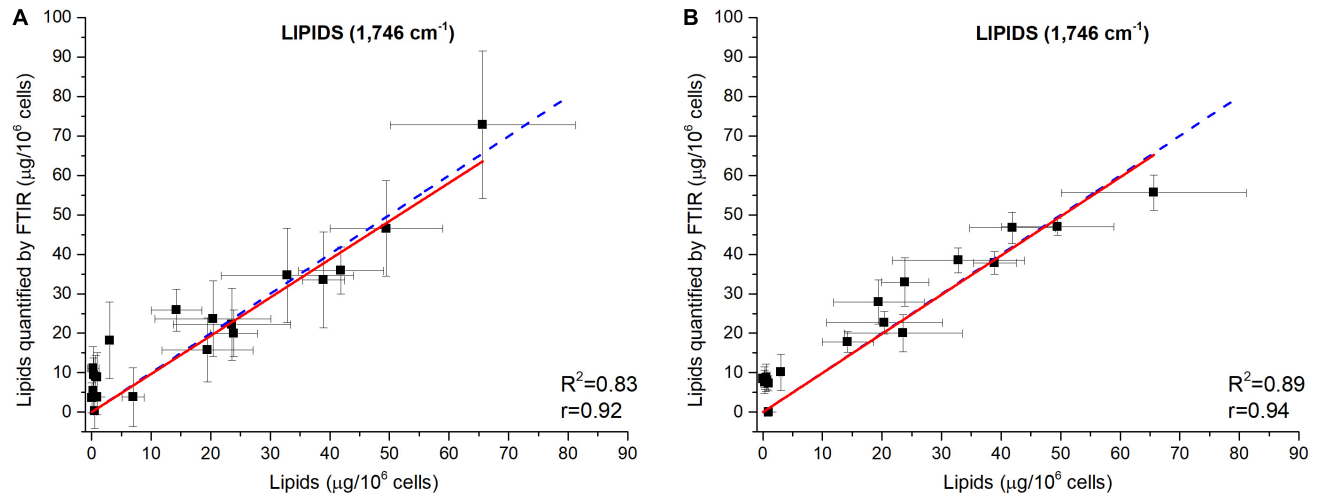


FIGURE 5 | Validation of the FTIR methodology using data from a N-starvation experiment conducted in a photobioreactor (Scarsini, 2021). Correlations between lipids quantified from FTIR absorbance measured at 1,746 cm^{-1} and lipids concentrations quantified by fluorescence after Nile Red staining considering **(A)** baseline-corrected spectra and **(B)** baseline-corrected and normalized spectra. Two million cells were deposited on the silica plate. Adjusted R^2 and Pearson's correlation coefficients of the linear regression curves (r) are indicated on the graphs.

determined manually (Malassez or Neubauer hemacytometers). Hence, normalization steps avoid determining cell density of the culture which is, in addition to error prone, time-consuming.

Correlations Between Fourier Transform Infrared Spectroscopy Absorbance and Biochemical Quantification of Metabolites

The power of FTIR spectroscopy is the quantification at one glance of the three major macromolecular pools (Q_{Lipids}, $Q_{Proteins}$ and $Q_{Carbohydrates}$). In literature, both external or internal references have been exploited for their quantification. For example, Wagner et al. (2010) used glucose, glycerol tripalmitate and bovine serum albumin as external references for quantification calibration. However, a biological sample is a very complex matrix containing a mix of molecules, the interactions of which are reflected in the vibrational spectrum (Larkin, 2011). Exploiting external references for the generation of calibration curves could introduce a bias since molecular interactions are not correctly reproduced. This work proposes a different approach: the use of biological samples of *P. tricornutum* as standards. To generate the different standards, a serial dilution of a N-starved Pt4 culture with a N-repleted Pt4 culture was performed to provide a large range of values for lipid, protein, and carbohydrate cell quotas. For each serial dilution, Q_{Lipids}, $Q_{Proteins}$, and $Q_{Carbohydrates}$ were determined through conventional biochemical assays. Having determined the effective cell quota and to be in the linear range of the Beer-Lambert's law, samples of 2×10^6 cells from the different serial dilutions were deposited on the silica plate to perform the FTIR analyses described above. Strong linear correlations were obtained between biochemically quantified lipids and the absorbance at 1,746 cm^{-1} either after baseline correction (**Figure 3A**) or normalization (**Figure 3B**). Correlations for the other wavenumbers are found in **Supplementary Figure 8** and regression equations after baseline correction and/or normalization indicated in **Table 2**.

The results show that the organic macromolecules classes (lipids, proteins, and carbohydrates) can be quantified exploiting the IR absorbance at their specific wavenumber after the generation of proper calibration curves (*i.e.* using biochemical quantification methods). Performing a normalization increases the reliability of the quantification, particularly for proteins.

Validation of the Fourier Transform Infrared Spectroscopy Methodology Using Data From a N-Starvation Experiment Conducted in a Photobioreactor

The ability of the FTIR methodology to give results similar to those obtained by traditional methods (Scarsini, 2021) has been tested. **Figure 4** shows how both methodologies are comparable. Linear fittings between FTIR quantifications and measured quotas were performed (**Figure 5** for lipids at 1,746 cm^{-1} and **Supplementary Figures 9**, **10** for carbohydrates, proteins and fatty acids). The blue dotted lines represent theoretical values that must be expected while black squares are the plotted data (biochemical *versus* FTIR quantifications). Linear regressions of the plotted data are reported as well (red continuous lines). Adjusted R^2 and Pearson's correlation coefficient of the linear regression curves (r) are higher when normalization is performed (**Figure 5A** compared to **Figure 5B** and **Supplementary Figure 10** compared to **Supplementary Figure 9**), confirming that normalizing spectra allows a better quantification. Moreover, the high values of R^2 (>0.83) and r

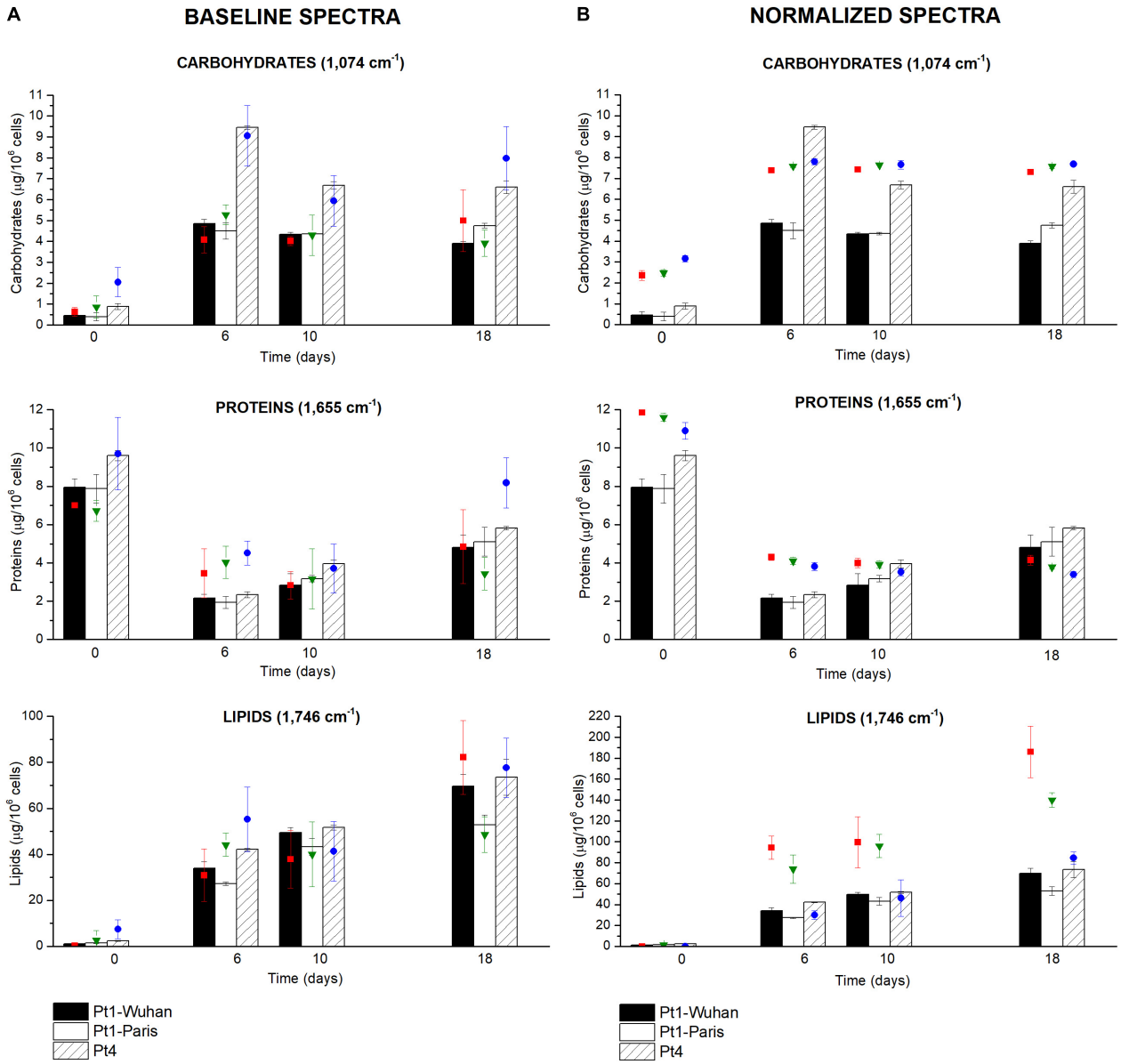


FIGURE 6 | Carbohydrates, proteins and lipids quantifications through the FTIR based methodology performed during a N-starvation of 18 days in the different ecotypes and strains of *P. tricornutum*. Quantifications performed with standard biochemical protocols are presented as histograms in the graph while quantifications using FTIR absorbance are represented as colored-dots (red square, green triangle, and blue dots), considering **(A)** baseline-corrected spectra or **(B)** baseline and normalized-corrected spectra. FTIR quantifications are performed at 1,074 cm^{-1} for carbohydrates, 1,655 cm^{-1} for proteins, and 1,746 cm^{-1} for lipids.

(>0.92) confirm the coherence between the two quantification methods (excepted for fatty acids wavenumbers).

The relatively lower R^2 ($0.83<R^2<0.89$) obtained for the wavenumber 1,746 cm^{-1} (**Figure 5**) could be due to the specificity of the biochemical quantification technique which is designed to quantify neutral lipids while FTIR quantify lipids regardless their class. Scarsini (2021) showed how in N-starvation, cells are principally accumulating neutral lipids through membrane lipid remodeling. The comparison of lipid content estimation through vibrational spectroscopy and a total lipids quantification through another biochemical technique, such as gravimetric analyses (Patel et al., 2019) could potentially reveal a direct linear correlation. The fatty acids measurements (2,855–2,954 cm^{-1}) do not show high correlations, especially when the normalization correction is applied ($R^2<0.43$ and $r<0.68$), probably due to the partial overlapping of these peaks with the large water peak at 3,000–3,700 cm^{-1}. Although better correlations are found for fatty acids with only a baseline correction ($R^2>0.76$ and $r>0.67$), the peak at 1,746 cm^{-1} is much more representative of the lipid accumulation and should thus be used for quantification. For proteins quantification, both 1,548 or 1,655 cm^{-1} can be used. Each of the five wavenumbers 1,043, 1,074, 1,107,

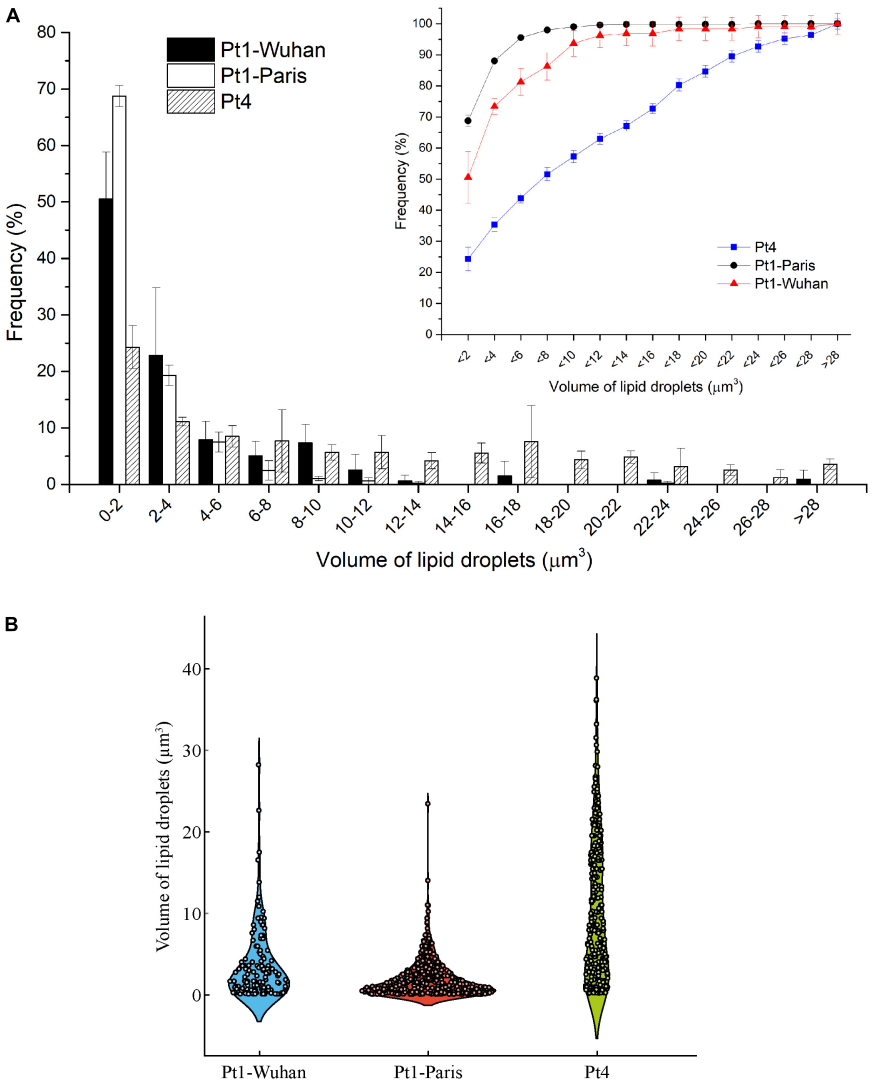


FIGURE 7 | Volume analysis of LDs in the different ecotypes and strains of *P. tricornutum* after 18 days of N-starvation. **(A)** Histogram reporting the occurrence frequencies for LDs of specific dimension ranges. Line graph reports the cumulative occurrence frequencies for increasing size of LDs; **(B)** Violin plot reporting the occurrence frequency distribution of the different size of LDs.

1,142, or 1,156 cm^{-1} can be used for a correct quantification of carbohydrates.

Application of Fourier Transform Infrared Spectroscopy Analysis to Infer Differences in Metabolites Quantifications Between Ecotypes of *P. tricornutum* During a N-Starvation Experiment

Around the world, ten different *P. tricornutum* ecotypes have been isolated and characterized. Being different at both genotypic and phenotypic levels, the ecotypes may also diverge in their cellular quota in major organic macromolecules (Martino et al., 2007; Rastogi et al., 2020). The described FTIR methodology was applied to characterize quantitatively the biochemical changes during 18 days of N-starvation exposure on two different ecotypes of *P. tricornutum i.e.* Pt1 and Pt4 with the aim to describe the potential metabolic divergences between them. **Figure 6** compares the different cellular quota determined using conventional biochemical quantifications (bars) with those obtained using FTIR measurements (colored-dots, carbohydrates: 1,074 cm^{-1}, proteins: 1,655 cm^{-1}, lipids: 1,746 cm^{-1}) after baseline correction (**Figure 6A**) and normalization (**Figure 6B**) using equations from **Table 2**.

Surprisingly, performing the normalization on spectra was not suitable for the comparison of metabolite contents between *P. tricornutum* ecotypes. The global trends of the different metabolites were similar (increase in carbohydrate and lipid contents and a decrease in protein content) but the differences between ecotypes were totally masked, indicating that the quantification using the SNV normalization may not be suitable

for comparing different microalgal species or ecotypes unless specific regression equations to each of them are applied. Much better correlations are found when spectra are only baseline-corrected, indicating that this data treatment may be used to compare different species or ecotypes, keeping in mind the greater variability of the data when spectra are not normalized.

Phenotypic Divergence Between *Phaeodactylum tricornutum* Ecotypes and Strains

Intrinsic differences between the two ecotypes can be observed in **Figure 6A**. Indeed, Pt4 shows a significantly higher $Q_{Carbohydrates}$ compared to both Pt1 all along the N-starvation: at day 6 of N-starvation, Pt4 contains twice as much $Q_{Carbohydrates}$ than both Pt1. The $Q_{Proteins}$ is also slightly higher, though not significant, for Pt4, especially in N-repleted condition (**Figure 6A**). No matter the ecotype considered, the Q_{Lipids} increased along the starvation, reflecting the reorganization of the carbon flux toward the production of storage lipids (*e.g.* Hockin et al., 2012; Levitan et al., 2015; McCarthy et al., 2017). The Q_{Lipids} of Pt4 18 days after the removal of N supply is significantly higher than that of Pt1-Paris. This result is confirmed by both FTIR spectroscopy and biochemical quantifications and strengthened by confocal microscopic analyses performed using BODIPY 505/515. Indeed, the distribution of LDs 10 days after the removal of N supply (**Figure 7**) indicates that LDs accumulated by Pt1-Paris are globally weaker than Pt1-Wuhan themselves smaller than Pt4. While Pt1-Paris seems to possess a single population of small LDs ($<2\ \mu m^3$), Pt4 possesses two to three subpopulations in the culture: the first one in the lowest ranges ($<2\ \mu m^3$) and the second and third ones peaking around 6–8 and 16–18 μm^3 (**Figure 7A**). This result is in line with the observations of Wong and Franz (2013) who found that Pt4 LDs ranged from 0.1 to 38 μm^3 in N-deficient conditions. Our results also tend to confirm what suggested Jaussaud et al. (2020) and Leyland et al. (2020) using epifluorescence microscopy. Indeed, Leyland et al. (2020) observed a higher yield of LDs from cells disrupted by hypotonic shock in Pt4 compared to Pt1 with multiple small LDs for Pt1 and two relatively large LDs for Pt4. By analyzing the dynamics of LD formation in Pt1, Jaussaud et al. (2020) highlighted the stepwise generation of subpopulations, at least three, marked by an increase in size over time reaching a maximum value. Here, we confirm the existence of LDs subpopulations but also the presence of larger LDs for Pt4 (**Figure 7B**). Jaussaud et al. (2020) suggested that LDs grow until they reach a size limited by cell packing with other membrane organelles and by the stiffness of the limiting cell wall. They also observed larger LDs in a mutant line having larger cells compared to wild type cells. The presence of larger LDs in Pt4 could be due to larger Pt4 cells compared to Pt1. Indeed, cell measurements performed on brightfield microscope pictures showed differences in cell dimensions between the different ecotypes 18 days after the removal of N supply (**Supplementary Figure 11**). In particular, Pt4 ecotype appears 1.5 times wider than Pt1 ecotypes while small or no differences are observed in length.

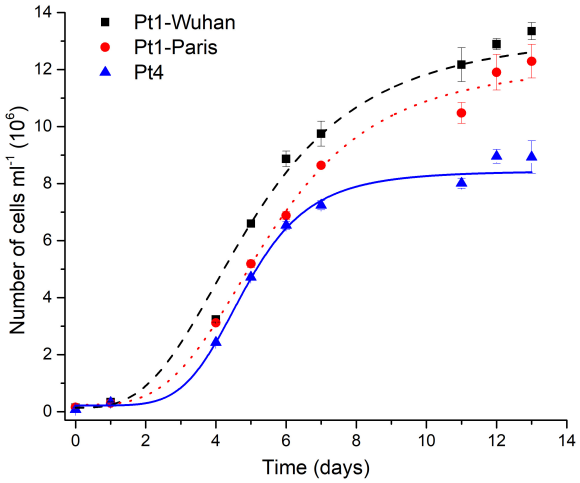


FIGURE 8 | Growth curves of the different ecotypes and strains of *P. tricornutum* in N-replete condition. A logistic model was used for the determination of growth rates.

TABLE 3 | Rules to follow for a suitable quantification of lipids, proteins and carbohydrates in *P. tricornutum* through the described FTIR technique.

Feature	Parameter
FTIR modus	Transmission
Parameters	4 cm^{-1} resolution, 32 scans/spectrum
Number of cells to deposit on the silica plate	1×10^6 to 3×10^6 cells
Number of biological replicates	3
Number of technical replicates	8
Collection of the raw spectrum	800–2,000 cm^{-1}
Baseline-correction	Rubber-band
Normalization	Standard normal variate
Lipids quantification	1,746 cm^{-1}
Proteins quantification	1,548 or 1,655 cm^{-1}
Carbohydrates quantification	1,043, 1,074, 1,107, 1,142, or 1,156 cm^{-1}

Pigment quantifications assessed (**Supplementary Figure 12**), revealed a higher total carotenoid (mainly represented by fucoxanthin) content and a lower Chl *a*/Chl c ratio in Pt4 than in Pt1, especially in N-repleted condition. According to Lamote et al. (2003), the lower values of the Chl *a*/Chl c ratio suggests a larger light harvesting antenna in Pt4 than in Pt1. This interpretation fits with the higher content in fucoxanthin in Pt4 since it is the main pigment of the fucoxanthin–protein complexes forming the light-harvesting antenna in diatoms (for a review see Büchel, 2020). This result is in line with the lower non-photochemical quenching capacity observed for Pt4 and proposed to be an adaptive trait to low light conditions (Bailleul et al., 2010). This accession has also been proposed to establish an upregulation of a peculiar light harvesting protein LHCX4 in extended dark conditions (Bailleul et al., 2010; Taddei et al., 2016). In N-repleted condition, a significantly higher maximal division rate was observed for Pt4 ($0.99 \pm 0.05\ day^{-1}$, $p < 0.05$)

compared to Pt1-Wuhan (0.73 ± 0.06 day^{-1}) and Pt1-Paris (0.67 ± 0.05 day^{-1}) but also a lower maximal cell density for Pt4 (8×10^6 of cells ml^{-1} while being around 11 and 12×10^6 of cells ml^{-1} for respectively Pt1-Paris and Pt1-Wuhan) (**Figure 8**). These results highlight a faster division rate and a higher N intake cell^{-1} for Pt4 compared to both Pt1 strains. In line with these observations, Rastogi et al. (2020) pointed out the presence of a higher copy number in the genome of Pt4 of a gene involved in nitrate assimilation (Phatr3_EG02286 encoding a nitrite reductase), suggesting an altered mode of nutrient acquisition in Pt4 compared to the other ecotypes.

Altogether, our results highlight that *P. tricornutum* ecotypes differ in their major macromolecules cell quotas and nutrient acquisition intensity: Pt4 cells absorb more nitrate and funnel more energy in the photosynthetic apparatus with consequently a higher ATP production and a faster division in N-repleted conditions. When facing N-starvation, Pt4 cells have a higher protein and carbohydrates content coupled with bigger cytoplasmic LDs, probably related to their larger cell volume compared to Pt1 cells. These results strengthen the genetic divergence already highlighted between Pt1 and Pt4. Indeed, using whole genome sequencing, Rastogi et al. (2020) draw a comprehensive landscape of the genomic diversity between the 10 accessions of *P. tricornutum*. They described strong genetic subdivisions of the accessions into four genetic clades with populations of each clade possessing a conserved genetic and functional makeup, likely a consequence of the limited dispersal of *P. tricornutum* in the open ocean. They observed that Pt4 was the most genetically divergent ecotype. However, the link between genetic and phenotypic divergence still remains to be elucidated.

Interestingly, the two Pt1 lines differ in their lipid content (**Figures 6**, **7**). Pt1-Wuhan seems to possess, in addition to the small population of LDs present in Pt1-Paris (<2 μm^3), a second subpopulation of LDs of higher volumes (8–10 μm^3) (**Figures 7A,B**). This finding strengthens the significantly higher accumulation of neutral lipids measured using Nile red fluorescence in Pt1-Wuhan compared to Pt1-Paris. Although both lines originated from the same Culture Collection (CCMP 2561), this result may highlight a different evolutionary adaptation of the two strains induced by the culture conditions under which both Pt1 strains were maintained. As reported by Lakeman et al. (2009), in laboratory cultures, because of rapid growth rates and high population densities, mutations affecting the phenotype are likely to arise. Different processes can lead to genetic evolution in strains maintained in different laboratories. Genetic drift is likely to occur when serial transfer of microalgae are performed for maintenance: a relatively small and random inoculum of the genetic diversity is taken from the parent culture to start a new batch, introducing potentially genetic drift into the evolutionary history of the strain (Lakeman et al., 2009). Selective pressure may also be exerted by culture conditions. It occurs every time culture conditions are altered (temperature, irradiance, salinity, nutrients, etc.) and can be as subtle as a shift in light quality when growth chamber bulbs are replaced or as profound as a modified growth medium when a strain is sent to a new laboratory. The more times a strain is transferred between laboratories and subjected to novel growth conditions, the greater the potential for it to evolve properties that deviate from those of its original phenotype (Lakeman et al., 2009).

CONCLUSION

Fourier transform infrared spectroscopy methodology developed in this study is suitable for the quantification of cellular quota of the major metabolites classes of unicellular organisms provided some simple rules are followed (**Table 3**): (1) deposit the adequate amount of cells, (2) performing baseline correction on spectra collected at 800–200 cm^{-1}, (3) performing the SNV normalization because it contributes to decrease the variability between biological samples, and (4) choose the adequate wavenumbers for biomolecule quantifications. In comparison to conventional chemical analyses, FTIR spectroscopy has striking advantages: (1) a high reliability and sensitivity, (2) a high speed of measurement procedure, (3) a low volume of sample is necessary, (4) no need to perform complex and time-consuming extractions, and (5) no need to determine the sample cell density which is time-consuming and subjected to error-prone. Data reported in this paper also confirm the metabolic divergence between Pt1 and Pt4 and even highlight phenotypic divergence between strains of the same ecotype cultured in laboratories with probably different maintenance modes. This suggests that exploring the intra-species processes of evolution can be trickier than expected and that even in our controlled laboratory environments, evolution do not cease to exert its influence.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

BS, JM, and MS: conceptualization, data curation, and initial writing. BS, JM, MS, AT, and MB: formal analysis. BS and JM: funding acquisition, project administration and supervision. BS, BV, JM, MS, and AT: investigation. BS, BV, JM, MS, FA, FL, FN, and AT: methodology. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

Authors are grateful to the IMMM platform of Le Mans University for the service provided for FTIR spectroscopy and confocal microscopy. Authors are also thankful to Angela Falciatore from the Laboratoire de Biologie du Chloroplaste et Perception de la Lumière chez les Microalgues, Institut de Biologie Physico-Chimique (IBPC, Paris, France) and to Hanhua Hu from the Center for Algal Biology and Applied Research (Chinese Academy of Sciences at Wuhan, China) for sharing

their *P. tricornutum* Pt1 strains. We are also grateful to Le Mans University for the Ph.D. grant to MS and the post doctorate position to AT. The "Fondation Banque Populaire de l'Ouest", Le Mans University and the University "Bretagne-Loire" (UBL) are thanked for financial supports.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.756421/full#supplementary-material

REFERENCES

Bailleul, B., Rogato, A., Martino, A., De, Coesel, S., Cardol, P., Bowler, C., et al. (2010). An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proc. Natl. Acad. Sci.* 107, 18214–18219. doi: 10.1073/pnas.1007703107

Bajhaiya, A. K., Dean, A. P., Driver, T., Trivedi, D. K., Rattray, N. J. W., Allwood, J. W., et al. (2015). High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation. *Metabolomics* 121, 1–14. doi: 10.1007/S11306-015-0878-4

Barnes, R. J., Dhanoa, M. S., and Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43, 772–777. doi: 10.1366/0003702894202201

Bowler, C., and Falciatore, A. (2019). Phaeodactylum tricornutum. *Trends Genet.* 35, 706–707. doi: 10.1016/j.tig.2019.05.007

Büchel, C. (2020). *Light-harvesting Complexes of Diatoms: fucoxanthin-Chlorophyll Proteins*. Germany: Springer. 441–457. doi: 10.1007/978-3-030-33397-3_16

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). *Phaeodactylum tricornutum*: a diatom cell factory. *Trends Biotechnol.* 38, 1–17. doi: 10.1016/j.tibtech.2019.12.023

Cointet, E., Wielgosz-Collin, G., Méléder, V., and Gonçalves, O. (2019). Lipids in benthic diatoms: a new suitable screening procedure. *Algal Res.* 39:101425. doi: 10.1016/J.ALGAL.2019.101425

Cooper, M. S., Hardin, W. R., Petersen, T. W., and Cattolico, R. A. (2010). Visualizing "green oil" in live algal cells. *J. Biosci. Bioeng.* 109, 198–201. doi: 10.1016/j.jbiosc.2009.08.004

Dean, A. P., Nicholson, J. M., and Sigee, D. C. (2008). Impact of phosphorus quota and growth phase on carbon allocation in *Chlamydomonas reinhardtii*: an FTIR microspectroscopy study. *Eur. J. Phycol.* 43, 345–354. doi: 10.1080/09670260801979287

Dean, A. P., Sigee, D. C., Estrada, B., and Pittman, J. K. (2010). Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. *Bioresour. Technol.* 101, 4499–4507. doi: 10.1016/j.biortech.2010.01.065

Driver, T., Bajhaiya, A. K., Allwood, J. W., Goodacre, R., Pittman, J. K., and Dean, A. P. (2015). Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal Res.* 11, 148–155. doi: 10.1016/j.algal.2015.06.009

Eid, K. F., Panth, M., and Sommers, A. D. (2018). The physics of water droplets on surfaces: exploring the effects of roughness and surface chemistry. *Eur. J. Phys.* 39:025804. doi: 10.1088/1361-6404/aa9cba

Giordano, M., Kansiz, M., Heraud, P., Beardall, J., Wood, B., and McNaughton, D. (2001). Fourier Transform Infrared spectroscopy as a novel tool to investigate changes in intracellular macromolecular pools in the marine microalga *Chaetoceros muellerii* (Bacillariophyceae). *J. Phycol.* 37, 271–279. doi: 10.1046/j.1529-8817.2001.037002271.x

Giordano, M., Norici, A., and Beardall, J. (2017). Impact of inhibitors of amino acid, protein, and RNA synthesis on C allocation in the diatom *Chaetoceros muellerii*: a FTIR approach. *Algae* 32, 161–170. doi: 10.4490/algae.2017.32.6.6

Gordon, R., Gordon, R., Merz, C. R., Gurke, S., and Schoefs, B. (2019). "Bubble farming: scalable microcosms for diatom biofuel and the next green revolution" in *Diatoms: fundamentals and Applications*. (Eds) M. A. Lapidge, J. Seckbach and R. Gordon. (United States: Wiley). 583–654. doi: 10.1002/9781119370741.ch22

Govender, T., Ramanna, L., Rawat, I., and Bux, F. (2012). BODIPY staining, an alternative to the Nile Red fluorescence method for the evaluation of intracellular lipids in microalgae. *Bioresour. Technol.* 114, 507–511. doi: 10.1016/j.biortech.2012.03.024

Guillard, R. R. L. (1975). "Culture of phytoplankton for feeding marine invertebrates," in *Culture of Marine Invertebrate Animals*. (eds). Smith W.L., Chanley M.H. (Boston: Springer). 29–60. doi: 10.1007/978-1-4615-8714-9_3

Heinemann, B., and Paulsen, H. (1999). Random mutations directed to transmembrane and loop domains of the light-harvesting chlorophyll a/b protein: impact on pigment binding. *Biochemistry* 38, 14088–14093. doi: 10.1021/bi991439a

Heraud, P., Wood, B. R., Tobin, M. J., Beardall, J., and McNaughton, D. (2005). Mapping of nutrient-induced biochemical changes in living algal cells using synchrotron infrared microspectroscopy. *FEMS Microbiol. Lett.* 249, 219–225. doi: 10.1016/j.femsle.2005.06.021

Heydarizadeh, P., Boureba, W., Zahedi, M., Huang, B., Moreau, B., Lukomska, E., et al. (2017). Response of CO2-starved diatom *Phaeodactylum tricornutum* to light intensity transition. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160396. doi: 10.1098/rstb.2016.0396

Heydarizadeh, P., Marchand, J., Chenais, B., Sabzalian, M. R., Zahedi, M., Moreau, B., et al. (2014). Functional investigations in diatoms need more than transcriptomic approach. *Diatom Res.* 29, 75–89. doi: 10.1080/0269249X.2014.883727

Heydarizadeh, P., Veidl, B., Huang, B., Lukomska, E., Wielgosz-Collin, G., Couzinet-Mossion, A., et al. (2019). Carbon orientation in the diatom *Phaeodactylum tricornutum*: the effects of carbon limitation and photon flux density. *Front. Plant Sci.* 10, 1–16. doi: 10.3389/fpls.2019.00471

Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 158, 299–312. doi: 10.1104/pp.111.184333

Hu, H., and Pan, Y. (2020). Electroporation transformation protocol for *Phaeodactylum tricornutum*. *Methods Mol. Biol.* 2050, 163–167. doi: 10.1007/978-1-4939-9740-4_17

Huang, B., Marchand, J., Blanckaert, V., Lukomska, E., Ulmann, L., Wielgosz-Collin, G., et al. (2019). Nitrogen and phosphorus limitations induce carbon partitioning and membrane lipid remodelling in the marine diatom *Phaeodactylum tricornutum*. *Eur. J. Phycol.* 54, 342–358. doi: 10.1080/09670262.2019.1567823

Jaussaud, A., Lupette, J., Salvaing, J., Jouhet, J., Bastien, O., Gromova, M., et al. (2020). Stepwise biogenesis of subpopulations of lipid droplets in nitrogen starved *Phaeodactylum tricornutum* cells. *Front. Plant Sci.* 11:48. doi: 10.3389/fpls.2020.00048

Lakeman, M. B., von Dassow, P., and Cattolico, R. A. (2009). The strain concept in phytoplankton ecology. *Harmful Algae* 8, 746–758. doi: 10.1016/j.hal.2008.11.011

Lamote, M., Darko, E., Schoefs, B., and Lemoine, Y. (2003). Assembly of the photosynthetic apparatus in embryos from Fucus serratus l. *Photosynth. Res.* 77, 45–52. doi: 10.1023/A:1024999024157

Larkin, P. (2011). Environmental dependence of vibrational spectra. *Infrared Raman Spectrosc.* 2011, 55–62. doi: 10.1016/b978-0-12-386984-5.10004-7

Levitan, O., Dinamarca, J., Zelzion, E., Lun, D. S., Guerra, L. T., Kim, M. K., et al. (2015). Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *Proc. Natl. Acad. Sci.* 112, 412–417. doi: 10.1073/pnas.1419818112

Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S., and Khozin-Goldberg, I. (2020). High resolution proteome of lipid droplets isolated from the pennate diatom *Phaeodactylum tricornutum* (Bacillariophyceae) strain pt4 provides mechanistic insights into complex intracellular coordination during nitrogen deprivation. *J. Phycol.* 56, 1642–1663. doi: 10.1111/jpy.13063

Mampallil, D. (2014). Some physics inside drying droplets. *Resonance* 19, 123–134. doi: 10.1007/s12045-014-0016-z

Martino, A., De, Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions [1]. *J. Phycol.* 43, 992–1009. doi: 10.1111/j.1529-8817.2007.00384.x

Mayers, J. J., Flynn, K. J., and Shields, R. J. (2013). Chalmers Publication Library Rapid determination of bulk microalgal biochemical composition by Fourier-Transform Infrared spectroscopy Rapid determination of bulk microalgal biochemical composition by Fourier-Transform Infrared spectroscopy. *Bioresour. Technol.* 148, 215–220. doi: 10.1016/j.biortech.2013.08.133

McCarthy, J. K., Smith, S. R., McCrow, J. P., Tan, M., Zheng, H., Beeri, K., et al. (2017). Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell.* 29, 2047–2070. doi: 10.1105/tpc.16.00910

Mimouni, V., Ulmann, L., Pasquet, V., Mathieu, M., Picot, L., Bougaran, G. G., et al. (2012). The potential of microalgae for the production of bioactive molecules of pharmaceutical interest. *Curr. Pharm. Biotechnol.* 13, 2733–2750. doi: 10.2174/138920112804724828

Movasaghi, Z., Rehman, S., and Rehman, I. U. (2008). Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* 43, 134–179. doi: 10.1080/05704920701829043

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Patel, A., Antonopoulou, I., Enman, J., Rova, U., Christakopoulos, P., and Matsakas, L. (2019). Lipids detection and quantification in oleaginous microorganisms: an overview of the current state of the art. *BMC Chem. Eng.* 1:13. doi: 10.1186/s42480-019-0013-9

Pogorzelec, N. M., Mundy, C. J., Findlay, C. R., Campbell, K., Diaz, A., Ehn, J. K., et al. (2017). FTIR imaging analysis of cell content in sea-ice diatom taxa during a spring bloom in the lower Northwest Passage of the Canadian Arctic. *Mar. Ecol. Prog. Ser.* 569, 77–88. doi: 10.3354/MEPS12088

Rastogi, A., Vieira, F. R. J., Deton-Cabanillas, A. F., Veluchamy, A., Cantrel, C., Wang, G., 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. doi: 10.1038/s41396-019-0528-3

Razak, N. F. A., Abd Karim, R. H., Jamal, J. A., and Said, M. M. (2020). Rapid discrimination of halal and non-halal pharmaceutical excipients by fourier transform infrared spectroscopy and chemometrics. *J. Pharm. Bioallied Sci.* 12, S752–S757. doi: 10.4103/jpbs.JPBS_364_19

Sandak, J., Sandak, A., and Allegretti, O. (2016). Chemical changes to woody polymers due to high-temperature thermal treatment assessed with near infrared spectroscopy. *J. Near Infrar. Spectros.* 24, 555–562. doi: 10.1255/jnirs.1220

Scarsini, M. (2021). *The Transition Kinetic Toward Nitrogen Deprivation in the Marine Diatom Phaeodactylum tricornutum: A Multidisciplinary Approach.* Ph.D. thesis. Le Mans University (France). Available online at: https://www.theses.fr/s206840

Scarsini, M., Marchand, J., Manoylov, K. M., and Schoefs, B. (2019). "Photosynthesis in diatoms" in *Diatoms: fundamentals and Applications*. eds J. Seckbach, and R. Gordon. (United States: John Wiley & Sons, Ltd). 191–211. doi: 10.1002/9781119370741.ch8

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9:3924. doi: 10.1038/s41467-018-06378-9

Shen, X., Ye, S., Xu, L., Hu, R., Jin, L., Xu, H., et al. (2018). Study on baseline correction methods for the Fourier transform infrared spectra with different signal-to-noise ratios. *Appl. Opt.* 57:5794. doi: 10.1364/ao.57.005794

Sigee, D. C., Bahrami, F., Estrada, B., Webster, R. E., and Dean, A. P. (2007). The influence of phosphorus availability on carbon allocation and P quota in Scenedesmus subspicatus: a synchrotron-based FTIR analysis. *Phycologia* 46, 583–592. doi: 10.2216/07-14.1

Stehfest, K., Toepel, J., and Wilhelm, C. (2005). The application of micro-FTIR spectroscopy to analyze nutrient stress-related changes in biomass composition of phytoplankton algae. *Plant Physiol. Biochem.* 43, 717–726. doi: 10.1016/j.plaphy.2005.07.001

Stukenberg, D., Zauner, S., Dell'Aquila, G., Maier, U. G., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 9:740. doi: 10.3389/fpls.2018.00740

Taddei, L., Stella, G. R., Rogato, A., Bailleul, B., Fortunato, A. E., Annunziata, R., et al. (2016). Multisignal control of expression of the LHCX protein family in the marine diatom *Phaeodactylum tricornutum*. *J. Exp. Bot.* 67, 3939–3951. doi: 10.1093/jxb/erw198

Van Wagenen, J., Holdt, S. L., De Francisci, D., Valverde-Pérez, B., Plósz, B. G., and Angelidaki, I. (2014). Microplate-based method for high-throughput screening of microalgae growth potential. *Bioresour. Technol.* 169, 566–572. doi: 10.1016/j.biortech.2014.06.096

Vinayak, V., Manoylov, K. M., Gateau, H., Blanckaert, V., Hérault, J., Pencréac'H, G., et al. (2015). Diatom milking? A review and new approaches. *Mar. Drugs* 13, 2629–2665. doi: 10.3390/md13052629

Wagner, H., Liu, Z., Langner, U., Stehfest, K., and Wilhelm, C. (2010). The use of FTIR spectroscopy to assess quantitative changes in the biochemical composition of microalgae. *J. Biophotonics* 3, 557–566. doi: 10.1002/jbio.201000019

Wong, D. M., and Franz, A. K. (2013). A comparison of lipid storage in *Phaeodactylum tricornutum* and *Tetraselmis suecica* using laser scanning confocal microscopy. *J. Microbiol. Methods* 95, 122–128. doi: 10.1016/j.mimet.2013.07.026

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Scarsini, Thurotte, Veidl, Amiard, Niepceron, Badawi, Lagarde, Schoefs and Marchand. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fpls.2021.779307

Global Profiling of N-Glycoproteins and N-Glycans in the Diatom *Phaeodactylum tricornutum*

Xihui Xie[1,2†], Hong Du[1,2†], Jichen Chen[1,2], Muhammad Aslam[1,2,3], Wanna Wang[1,2], Weizhou Chen[1,2], Ping Li[1], Hua Du[1,2] and Xiaojuan Liu[1,2]*

[1] Guangdong Provincial Key Laboratory of Marine Biotechnology, STU-UNIVPM Joint Algal Research Center, College of Sciences, Institute of Marine Sciences, Shantou University, Shantou, China, [2] Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou, China, [3] Faculty of Marine Sciences, Lasbela University of Agriculture, Water & Marine Sciences, Uthal, Pakistan

N-glycosylation is an important posttranslational modification in all eukaryotes, but little is known about the N-glycoproteins and N-glycans in microalgae. Here, N-glycoproteomic and N-glycomic approaches were used to unveil the N-glycoproteins and N-glycans in the model diatom *Phaeodactylum tricornutum*. In total, 863 different N-glycopeptides corresponding to 639 N-glycoproteins were identified from *P. tricornutum*. These N-glycoproteins participated in a variety of important metabolic pathways in *P. tricornutum*. Twelve proteins participating in the N-glycosylation pathway were identified as N-glycoproteins, indicating that the N-glycosylation of these proteins might be important for the protein N-glycosylation pathway. Subsequently, 69 N-glycans corresponding to 59 N-glycoproteins were identified and classified into high mannose and hybrid type N-glycans. High mannose type N-glycans contained four different classes, such as Man-5, Man-7, Man-9, and Man-10 with a terminal glucose residue. Hybrid type N-glycan harbored Man-4 with a terminal GlcNAc residue. The identification of N-glycosylation on nascent proteins expanded our understanding of this modification at a N-glycoproteomic scale, the analysis of N-glycan structures updated the N-glycan database in microalgae. The results obtained from this study facilitate the elucidation of the precise function of these N-glycoproteins and are beneficial for future designing the microalga to produce the functional humanized biopharmaceutical N-glycoproteins for the clinical therapeutics.

Keywords: N-glycosylation modification, N-glycoprotein, N-glycan, *Phaeodactylum tricornutum*, biopharmaceuticals

OPEN ACCESS

Edited by:
Hanhua Hu,
Institute of Hydrobiology, Chinese Academy of Sciences (CAS), China

Reviewed by:
Zhuo Chen,
Shandong Normal University, China
Tore Brembu,
Norwegian University of Science and Technology, Norway

***Correspondence:**
Xiaojuan Liu
liuxiaojuan@stu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *18 September 2021*
Accepted: *05 November 2021*
Published: *03 December 2021*

Citation:
Xie X, Du H, Chen J, Aslam M, Wang W, Chen W, Li P, Du H and Liu X (2021) Global Profiling of N-Glycoproteins and N-Glycans in the Diatom Phaeodactylum tricornutum. Front. Plant Sci. 12:779307. doi: 10.3389/fpls.2021.779307

INTRODUCTION

Diatoms are a group of unicellular eukaryotic microalgae and contribute approximately 32% of global phytoplankton primary production (Uitz et al., 2010). *Phaeodactylum tricornutum*, a pennate diatom, is an ideal model organism for basic research of biotechnology, due to availability of its complete sequenced genome (Bowler et al., 2008) and convenient genetic operating systems (e.g., established overexpression, knock-down, and knock-out of target genes) (Daboussi et al., 2014; Weyman et al., 2015; Nymark et al., 2016). In addition, *P. tricornutum* has been

used as an alternative model organism for the expression of recombinant biopharmaceutical N-glycoproteins (Hempel et al., 2017; Vanier et al., 2018). Especially, a recombinant monoclonal antibody directed against the Hepatitis B virus surface antigen was reported to be modified by high mannose type N-glycans (Vanier et al., 2018). However, the N-glycosylation pathway in *P. tricornutum*, a major post-translational modification in the maturation of proteins, has received very little attention. Mathieu-rivet et al. (2013) demonstrated that protein N-glycosylation in *Chlamydomonas reinhardtii* occurs in the endoplasmic reticulum (ER) and Golgi apparatus. The initial steps of N-glycosylation pathway in the ER are highly conserved among almost all the eukaryotes. An oligosaccharide intermediate ($Man_5GlcNAc_2$-PP-Dol) first synthesizes on the cytosolic side of the ER, and then flipped into the lumen of ER, and continued the formation of oligosaccharide lipid precursor $Man_9GlcNAc_2$-PP-Dol. Afterward, the precursor transfers onto asparagine residue of the conserved Asn-X-Ser/Thr/Cys motifs (X is not proline). The newly synthesized glycoproteins are then recognized by a well-known ER quality control mechanism (ERQC) for their correct conformation (Blanco-Herrera et al., 2015). After one or multiple rounds of quality control, correct folded glycoproteins are exported from the ER to the Golgi apparatus for final mature modifications. During this ERQC mechanism, the unfolded/mis-folded glycoproteins are removed and then further processed/degraded by the ER-associated degradation pathway (ERAD) (Shin et al., 2018). The following steps of protein N-glycosylation in the Golgi apparatus are significantly different, finally giving rise to eukaryotic microalgae-specific N-glycans (Strasser, 2016). These N-glycans are classified into high mannose, paucimannose, complex, and hybrid types (Mathieu-Rivet et al., 2020).

Although the N-glycosylation pathway of proteins and mature N-glycan structures in yeasts, fungi and humans have been extensively studied. However, the information regarding the pathway of protein N-glycosylation in microalgae is still lacking. To date, the major mature N-glycans have been studied in green algae, red algae, and diatoms. For example, a complex type N-glycan bearing a core xylose residue was found in *Volvox carteri*, while the complex and oligomannosidic N-glycans were reported in *Scherffelia dubia* and *Tetraselmis striata* (Mathieu-Rivet et al., 2020). Moreover, N-glycans found in green microalga *Chlorella vulgaris* are found to be oligomannosidic N-glycans with 3-O-methyl and 3,6-di-O-methylmannose on the non-reducing terminus (Mócsai et al., 2019). The analysis of N-glycans in *Chlorella sorokiniana* demonstrated a huge heterogeneity and the presence of arabinose residues (Mócsai et al., 2020). In *Botryococcus braunii*, the synthesis of N-glycans was found to be dependent on N-acetylglucosaminyltransferase I (GnTI). Besides, it was found that N-glycans in *B. braunii* were modified by a N-acetylglucosamine (GlcNAc) residue at the non-reducing end and the (di)methylation of mannose residues (Schulze et al., 2017). The N-glycans in *C. reinhardtii* were analyzed *via* glycomic and glycoproteomic techniques which revealed that the synthesis of N-glycans is a GnTI-independent process (Mathieu-rivet et al., 2013). Moreover, oligomannosidic N-glycans (2–5 mannose residues) with 6-O-methylated mannoses and xylose residues are found to be main type in *C. reinhardtii* (Mathieu-rivet et al., 2013; Schulze et al., 2018). Later, the reinvestigation of N-glycans in *C. reinhardtii* showed that proteins carry linear $Man_5GlcNAc_2$ instead of the previous reported branched structure. It was speculated that the synthesis of linear N-glycans is owing to the lack of asparagine-linked glycosyltransferase (ALG)3, ALG6, and ALG12 activity (Vanier et al., 2017). In addition to green microalgae, O-methylated oligomannosidic N-glycan carrying xylose residues are identified from a cell wall glycoprotein of the red alga *Porphyridium* sp. (Levy-Ontman et al., 2011). High mannose type N-glycans with putative aminoethylphosphonate moieties are observed in euglenozoa *Euglena gracilis* (O'Neill et al., 2017). Moreover, it is reported that *P. tricornutum* proteins carries mainly high mannose type N-glycans (ranging from Man-5 to Man-9) in a GnTI-dependent pathway (Baïet et al., 2011). In addition to the bioinformatic analysis of genes involved N-glycosylation pathway, the reported N-glycosylation pathways in *C. reinhardtii* and *P. tricornutum* were updated and compared with the mammals in our previous study (Levy-Ontman et al., 2014; Liu et al., 2021). This comparative study revealed that N-glycosylation pathway of proteins and N-glycan structures are species-dependent. To get more reliable data, high throughput N-glycoproteomics and N-glycomics analysis have been considered to study the N-glycoproteins and N-glycans in *C. reinhardtii* (Mathieu-rivet et al., 2013), *B. braunii* (Schulze et al., 2017), and *Thalassiosira oceanica* (Behnke et al., 2021). However, the information regarding the presence of total N-glycoproteins and N-glycans in the model diatom, i.e., *P. tricornutum* still needs to be gathered.

In this study, *P. tricornutum* was used as a model organism to study the global profiling of N-glycoproteins and N-glycans using N-glycoproteomic and N-glycomic based approaches. We identified 863 N-glycopeptides in total from 639 glycoproteins in *P. tricornutum,* these proteins are involved in different cellular functions. Moreover, we identified 69 N-glycans corresponding to 59 N-glycoproteins. High mannose and hybrid type N-glycans were observed in this study. Identification of N-glycoproteins and N-glycans in *P. tricornutum* will not only deepen our understanding regarding the modification at N-glycoproteomic and N-glycomic scales but also facilitate the elucidation of the precise functions of these glycoproteins. Since it is well established that N-glycosylation affects the function and immunogenicity of recombinant pharmaceutical proteins. Therefore, this study is critically important for the expression of functional pharmaceutical N-glycoproteins in *P. tricornutum*.

MATERIALS AND METHODS

Microalga and Growth Conditions

Phaeodactylum tricornutum Pt1 (Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, China) cells were cultured in Erlenmeyer flasks containing f/2 medium constantly shaken at 225 rpm at 22°C under 24 h light condition (50 μmol photons m^{-2} s^{-1}). Algal cells were cultured for 5 days to

exponential phase in the f/2 medium. Subsequently, about 500 mg fresh algal cells were collected by centrifugation (1,500 × *g*, 10 min) for the next experiments.

Protein Extraction

Algal cells were grinded by liquid nitrogen into cell powder and transferred into 5 ml Eppendorf tube. Subsequently, lysis buffer, including 8 M urea, 1% Triton-100, 10 mM dithiothreitol, and 1% protease inhibitor cocktail (PIC), was added to the algal powder in a ratio of 1:4. The mixture was then sonicated 3–4 times on ice using a high intensity ultrasonic processor (Scientz). After the sonication, the debris was discarded by centrifugation at 20,000 × *g*, 4°C for 10 min. The protein from supernatant was precipitated with cold 20% TCA at −20°C for 2 h. The supernatant was removed after the centrifugation at 12,000 × *g*, 4°C for 10 min. The precipitation was washed three times using the cold acetone and then dissolved in 8 M urea. The protein concentration was measured with BCA kit according to the instructions of manufacturer (ab102536, Abcam, China) (Du et al., 2021).

Trypsin Digestion

For digestion, 5 mM dithiothreitol was used to reduce the protein solution at 56°C for 30 min and 11 mM iodoacetamide was used for the alkylation of protein at room temperature for 15 min in darkness. The protein sample was then diluted by adding 100 mM TEAB to urea concentration less than 2 M. Finally, trypsin was added for the first digestion overnight at 1:50 (trypsin-to-protein) mass ratio and for a second 4 h-digestion at 1:100 (trypsin-to-protein) mass ratio (Hockin et al., 2012).

High-Performance Liquid Chromatography Fractionation

The tryptic peptides were fractionated into fractions by high pH reverse-phase high-performance liquid chromatography (HPLC) using Thermo BetaSil C18 column (5 μm particles, 10 mm ID, 250 mm length). Briefly, the peptides were first separated with a gradient of 8–32% acetonitrile (pH 9.0) over 60 min into 60 fractions. Then, the peptides were combined into four fractions and dried by vacuum centrifuging (Kajiura et al., 2012).

Enrichment of Glycopeptides

The peptide mixtures were first dissolved in a loading buffer (80% acetonitrile/1% trifluoroacetic acid). The supernatant was transferred to hydrophilic interaction liquid chromatography (HILIC) microcolumn and centrifuged at 4,000 × *g* for 15 min. After three times washing with an enrichment buffer, the glycopeptides were collected from HILIC microcolumn by 10% acetonitrile and dried by vacuum freezing (Xiao et al., 2020). After drying, the glycopeptides were dissolved and digested in 50 μl 50 mM NH_4CO_3 buffer and 2 μl PNGase F at 37°C overnight. The digested glycopeptides were washed by C18 ZipTips and lyophilized for liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.

N-Glycoprotein Analysis *via* Liquid Chromatography Tandem Mass Spectrometry

The tryptic peptides were dissolved in 0.1% formic acid, directly loaded onto a home-made reversed-phase analytical column (15-cm length, 75 μm i.d.). The gradient was comprised of solvent A (0.1% formic acid in 2% acetonitrile), and an increase from 9 to 25% solvent B (0.1% formic acid in 90% acetonitrile) at 0–24 min, 25 to 35% solvent B in 24–32 min, and climbing to 35 to 80% in 33–36 min then, holding at 80% for the last 4 min (36–40 min), all at a constant flow rate of 350 nl/min on an EASY-nLC 1000 UPLC system (O'Neill et al., 2017).

The peptides were subjected to nano spray ionization (NSI) source followed by MS/MS in Q Exactive™ Plus (Thermo Fisher Scientific, Waltham, MA, United States) coupled online to the ultra-performance liquid chromatography (UPLC). The electrospray voltage applied was 2.0 kV. The m/z scan range was 350–1,800 for full scan, and intact peptides were detected in the Orbitrap at a resolution of 70,000. The peptides were then selected for MS/MS using normalized collision energy (NCE) setting as 28 and the fragments were detected in the Orbitrap at a resolution of 17,500. A data-dependent procedure that alternated between one MS scan followed by 20 MS/MS scans with 15.0 s dynamic exclusion. Automatic gain control (AGC) was set at 5E4. Signal threshold was set at 5,000 ions/s, the maximum injection time was set at 200 ms. Fixed first mass was set as 100 m/z.

N-Glycoprotein Database Search

The resulting MS/MS data were processed using MaxQuant search engine (v.1.5.2.8)[1]. The version of database is Uniprot_*Phaeodactylum_tricornutum*_strain_CCAP_1055 (10,465 entries). Tandem mass spectra were searched against *P. tricornutum* protein database concatenated with reverse decoy database. Trypsin/P was specified as cleavage enzyme allowing up to two missing cleavages. The mass tolerance for precursor ions was set as 20 ppm in the first search and 5 ppm in the main search, and the mass tolerance for fragment ions was set as 0.02 Da. Carbamidomethyl on Cys was specified as fixed modification and N-acetylated modification and Deamidation ^{18}O (N), Deamidation (NQ) were specified as variable modifications. False discovery rate (FDR) was adjusted to <1% and minimum score for modified peptides was set at >40. Motif analysis was carried out by MoMo (v5.0.2, motif-x algorithm)[2] (Cheng et al., 2019). And all the database protein sequences were used as background database parameter. Minimum number of occurrences was set to 20. Emulate original motif-x was ticked, and other parameters with default. Additionally, the annotations of GO and Domain were performed *via* InterProScan (v.5.14-53.0)[3] (Blum et al., 2021), and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation were performed by using KAAS (v.2.0)[4] and KEGG

[1]http://www.maxquant.org/

[2]http://meme-suite.org/tools/momo

[3]http://www.ebi.ac.uk/interpro/

[4]http://www.genome.jp/kaas-bin/kaas_main

Mapper (V2.5)[5] (Moriya et al., 2007; Kanehisa and Sato, 2020). Subcellular localization was predicted by several programs, such as SignalP 5.0[6], TargetP 2.0[7], ASAFind[8], Cell-PLoc 2.0[9], HECTAR v1.3[10], and WoLF PSORT[11]. An enrichment analysis was carried out through the Perl module (v.1.31)[12]. The heatmap of enrichment was predicted *via* R Package pheatmap (v.2.0.3)[13].

All differentially expressed modified protein database accession or sequences were searched against the STRING database version 10.5 for protein–protein interactions (PPI) (Szklarczyk et al., 2017). Only interactions between the proteins belonging to the searched data set were selected, thereby excluding the external candidates. STRING defines a metric called "confidence score" to define interaction confidence. We fetched all interactions that had a confidence score >0.7 (high confidence). Interaction network form STRING was visualized in R package "networkD3" (v.0.4)[14].

The data will be available online[15], the private project number is PXD022483.

N-Glycans Analysis *via* C18-Reversed Phase Liquid Chromatography Tandem Mass Spectrometry

About 2 g fresh microalgal cells were used to protein extraction, denaturation, reduction, alkylation, and digestion with trypsin. The digested peptides were then desalted using homemade C18 SPE tips and dried in a SpeedVac. Intact N-glycopeptides were enriched using ZIC-HILIC SPE tips (Liu et al., 2017; Xue et al., 2018). The enriched intact N-glycopeptides were subsequently reductively di-ethylated with acetaldehyde and acetaldehyde-$^{13}C_2$ (Fang et al., 2016). The isotopic labeled intact N-glycopeptides were desalted by a C18 SPE column, dried in the SpeedVac, and dissolved with ultrapure water (18.4 MΩ cm, Millipore Simplicity System) for further C18-RPLC-MS/MS (HCD) analysis.

Mass Spectrometry spectra were acquired in the m/z range 700–2,000 with mass resolution 60 k (m/z 200). The automatic gain control (AGC) target value and maximum injection time were placed at 3×10^6 and 20 ms. For MS/MS spectra, the mass resolution was set at 30 k. Fragmentation was obtained in a data-dependent acquisition (DDA) top 20 using HCD with stepped NCE (20, 30, and 31%). The AGC target value and maximum injection time were placed at 5×10^5 and 250 ms. Isolation window and dynamic exclusion were set at 3.0 m/z and 20.0 s.

[5]http://www.kegg.jp/kegg/mapper.html

[6]http://www.cbs.dtu.dk/services/SignalP/

[7]http://www.cbs.dtu.dk/services/TargetP/

[8]https://rocaplab.ocean.washington.edu/tools/asafind/

[9]http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/

[10]https://webtools.sb-roscoff.fr/?tool_id=toolsheddev.sb-roscoff.fr%2Frepos%2Flgueguen%2Fhectar_3_0%2Fabims_hectar%2F1.2&version=1.2&__identifer=vx6suh9fhag

[11]https://wolfpsort.hgc.jp/

[12]https://metacpan.org/pod/Text:NSP:Measures:2D:Fisher

[13]https://cran.r-project.org/web/packages/cluster/

[14]https://cran.r-project.org/web/packages/networkD3/

[15]http://beta.uniprot.org/

RESULTS

Identification of N-Glycoproteins in *P. tricornutum*

To acquire a global view of N-glycosylated proteins in *P. tricornutum*, total proteins were extracted and then digested with trypsin and fractionated by HPLC. The enriched N-glycopeptides were subsequently analyzed by LC-MS/MS with MaxQuant to classify the N-glycopeptide sequence with a maximum false discovery rate of 1% (**Figure 1A**). In total, 863 different N-glycosylated peptides corresponding to 639 proteins were identified. The detailed information is provided in **Supplementary Tables 1**, **2**. The mass errors of all the detected glycopeptides were distributed between −5 and +5 ppm, indicating that the mass accuracy of the MS data was fitting for the requirement of further analyses (**Figure 1B**). The length of the most identified N-glycopeptides was in the range of 9–23 amino acids, suggesting that the preparation of samples was standard (**Figure 1C**). Furthermore, it was shown that >60% of N-glycoproteins were modified at only one N-glycosylated site, few at two or multiple N-glycosylated sites (**Figure 1D**).

Motif Characters of the N-Glycopeptides

To further investigate the sequence models involving specific amino acids in all N-glycopeptides in *P. tricornutum*, flanking amino acid residues from position −10 to +10 around the N-glycosylated asparagine were analyzed by using a heatmap. The identified N-glycosylated peptides were analyzed in motif-X software to extract the overrepresented motifs of amino acids (**Figure 2**). Finally, four N-glycosylated motifs were obtained with the stringent significance. Strong bias in amino acids of specific N-glycosylated site motifs is observed. The sequence logos revealed that threonine (T) and Serine (S) are significantly overrepresented at +2 position (motif score >16), while glycine (G) and Serine (S) at +1 position (motif score >9.0) (**Figures 2A,B**). All the three amino acids are polar uncharged. For the identified 863 N-glycopeptides, 57.9% of the total N-glycosylation sites matched N-X-T (where X is a residue other than proline), with such sites being more abundant than those matching N-X-S (29.3%), N-G (6.8%), and N-S (6.0%). The modified site feature sequence and its detailed enrichment statistics are shown in **Table 1**.

Functional Annotation and Enrichment of N-Glycoproteins

Furthermore, the identified N-glycoproteins were all annotated by Clusters of Orthologous Group/Eukaryote Orthologous Group (COG/KOG) database (**Figure 3A**; Zheng et al., 2020). A total of 285 N-glycosylated proteins were annotated in COG/KOG database. These identified N-glycoproteins were mainly classified into five groups. 45 N-glycoproteins (15.79%) in posttranslational modification, protein turnover, and chaperones were identified. In total, 34 (11.93%) N-glycoproteins were annotated to be general function prediction only. About 30 (10.53%) N-glycoproteins were involved in carbohydrate transport and metabolism,

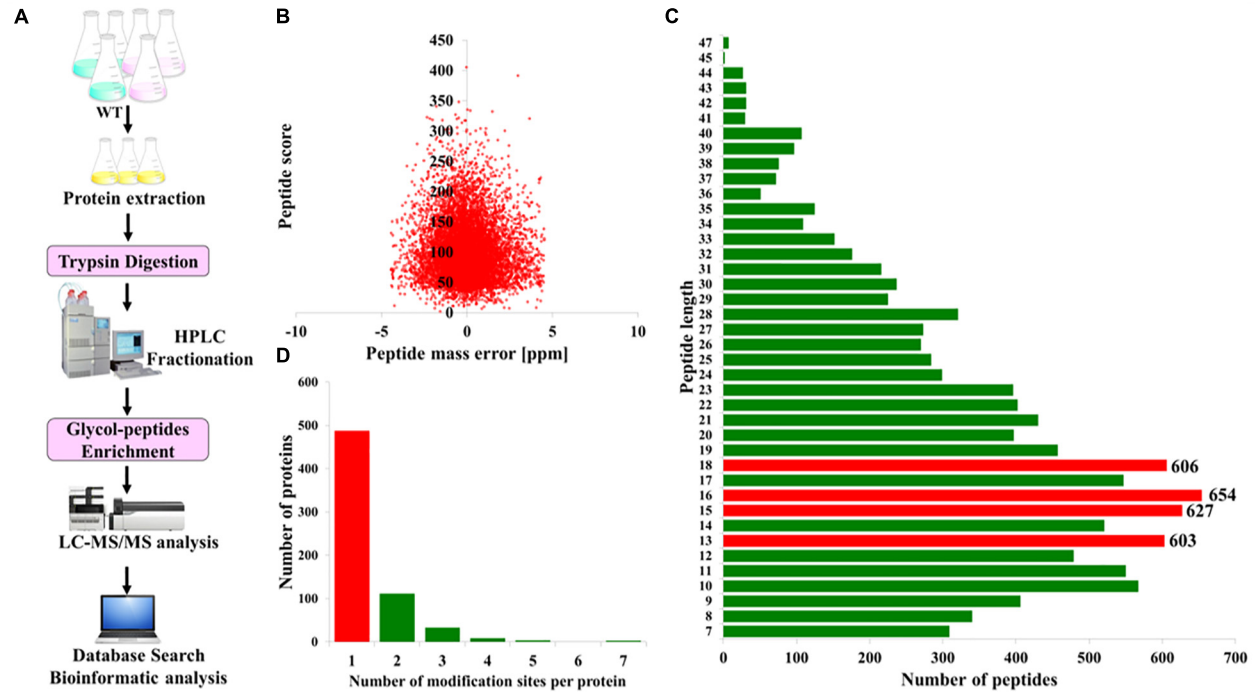


FIGURE 1 | Overview of the global identification of N-glycoproteins in *Phaeodactylum tricornutum.* **(A)** General workflow of the N-glycoprotein analysis. **(B)** Mass error distribution of the identified N-glycopeptides. **(C)** Length distribution of the N-glycopeptides. **(D)** The number of N-glycosylation modification sites per protein.

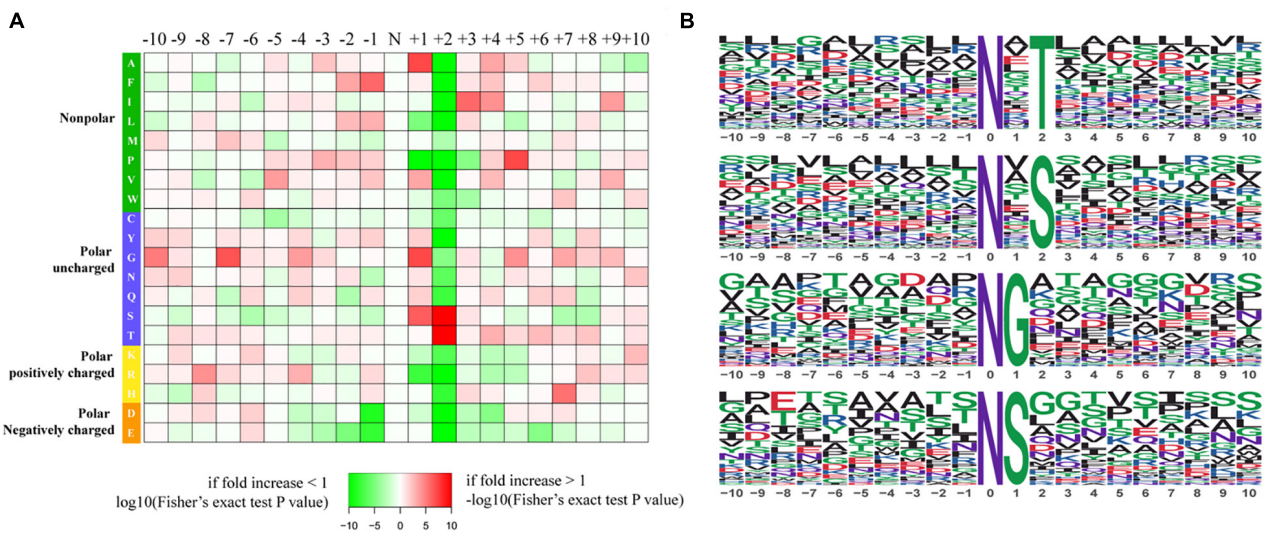


FIGURE 2 | Characteristics of the N-glycosylation sites. **(A)** The motif enrichment heatmap of upstream and downstream amino acids of all identified modification sites. Red indicates that this amino acid is significantly enriched near the modification site, and green indicates that this amino acid is significantly reduced near the modification site. **(B)** The main motifs modified by N-glycans.

and 26 (9.12%) N-glycoproteins were related with amino acid transport and metabolism. Additionally, 24 (8.42%) N-glycoproteins were identified into the energy production and conversion class (**Figure 3A**). The gene ontology (GO) analysis was also carried out to gain an insight into more implications of these N-glycoproteins in *P. tricornutum*. All N-glycosylated proteins were annotated and classified into three categories, such as biological process, cellular component, and molecular function (**Figure 3B**). In the biological process category, the most prevalent GO terms were metabolic process (231 N-glycoproteins), single-organism process (152 N-glycoproteins), and cellular process (144 N-glycoproteins)

TABLE 1 | The feature sequences of modified sites.

Motif	Motif score	Foreground		Background		Fold increase
		Matches	Size	Matches	Size	
xxxxxxxxxx_N_xTxxxxxxxx	16.00	425	893	10867	168907	7.4
xxxxxxxxxx_N_xSxxxxxxxx	16.00	215	468	14090	158040	5.2
xxxxxxxxxx_N_Gxxxxxxxxx	10.12	50	253	10380	143950	2.7
xxxxxxxxxx_N_Sxxxxxxxxx	9.03	44	203	10622	133570	2.7

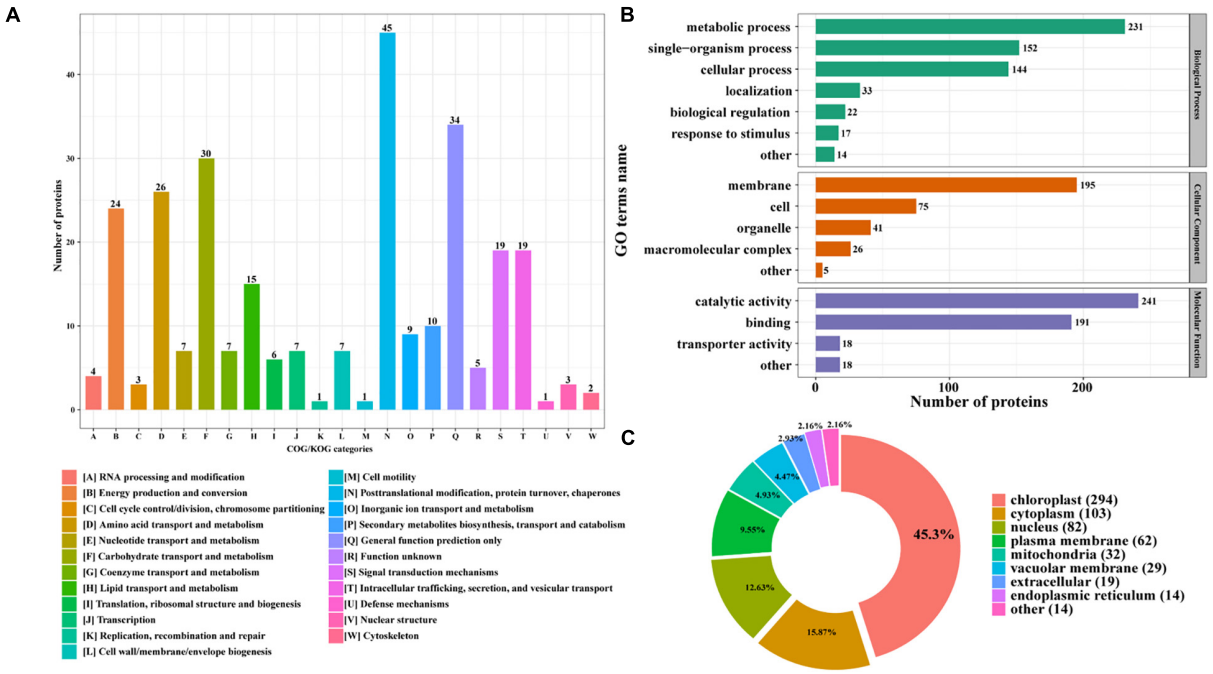


FIGURE 3 | Functional distribution and subcellular localization of the N-glycoproteins. **(A)** Clusters of Orthologous Group/Eukaryote Orthologous Group (COG/KOG) functional classification chart of N-glycoproteins. **(B)** Statistical distribution chart of N-glycoproteins under each GO category (second level). **(C)** Subcellular localization chart of N-glycoproteins.

(**Figure 3B**). In the cellular component, 195 N-glycoproteins were related with membrane. In the molecular function class, it is found that most N-glycosylated proteins are involved in the catalytic activity and binding.

The identified N-glycoproteins were classified based on subcellular localization, as shown in **Figure 3C**. According to the annotation, it is shown that 639 N-glycoproteins were putatively located to eight subcellular organelles, such as chloroplast, cytoplasm, nucleus, plasma membrane, mitochondria, vacuolar membrane, extracellular, and ER. Among them, most N-glycoproteins were targeted to chloroplast (294 N-glycoproteins) and cytoplasm (103 N-glycoproteins) (**Figure 3C**).

To further determine the proteins more prone to be N-glycosylation, the N-glycoproteins were enriched against all proteins from *P. tricornutum* genome database. The enrichment result of N-glycoproteins based on GO annotation is shown in **Figure 4**. All N-glycoproteins were annotated to three categories, molecular function, cellular component, and biological process. In molecular function category, most N-glycoproteins were enriched in hydrolase activity and acting on glycosyl bonds group. In cellular component category, glycosylphosphatidylinositol (GPI)-anchor transamidase complex related proteins were more prone to be N-glycosylation. While in biological process category, the proteins involved in monosaccharide metabolic process and hexose metabolic process were thought to be easier N-glycosylated. The analysis of KEGG pathway and protein domain enrichments are shown in **Figure 5**. In total, 171 proteins were identified to be involved in the KEGG pathways. The most abundant KEGG pathways were carbon and nitrogen metabolic pathways with 80 N-glycoproteins, accounting for 47% of all enriched proteins (**Figure 5A**). Besides, 175 proteins were enriched based on protein domains (**Figure 5B**). It was observed that proteins containing alpha/beta hydrolase fold (25, 14%), FAD/NAD(P)-binding domain (17, 10%), and glycoside hydrolase superfamily (13, 7%) domains were preferred to be N-glycosylated.

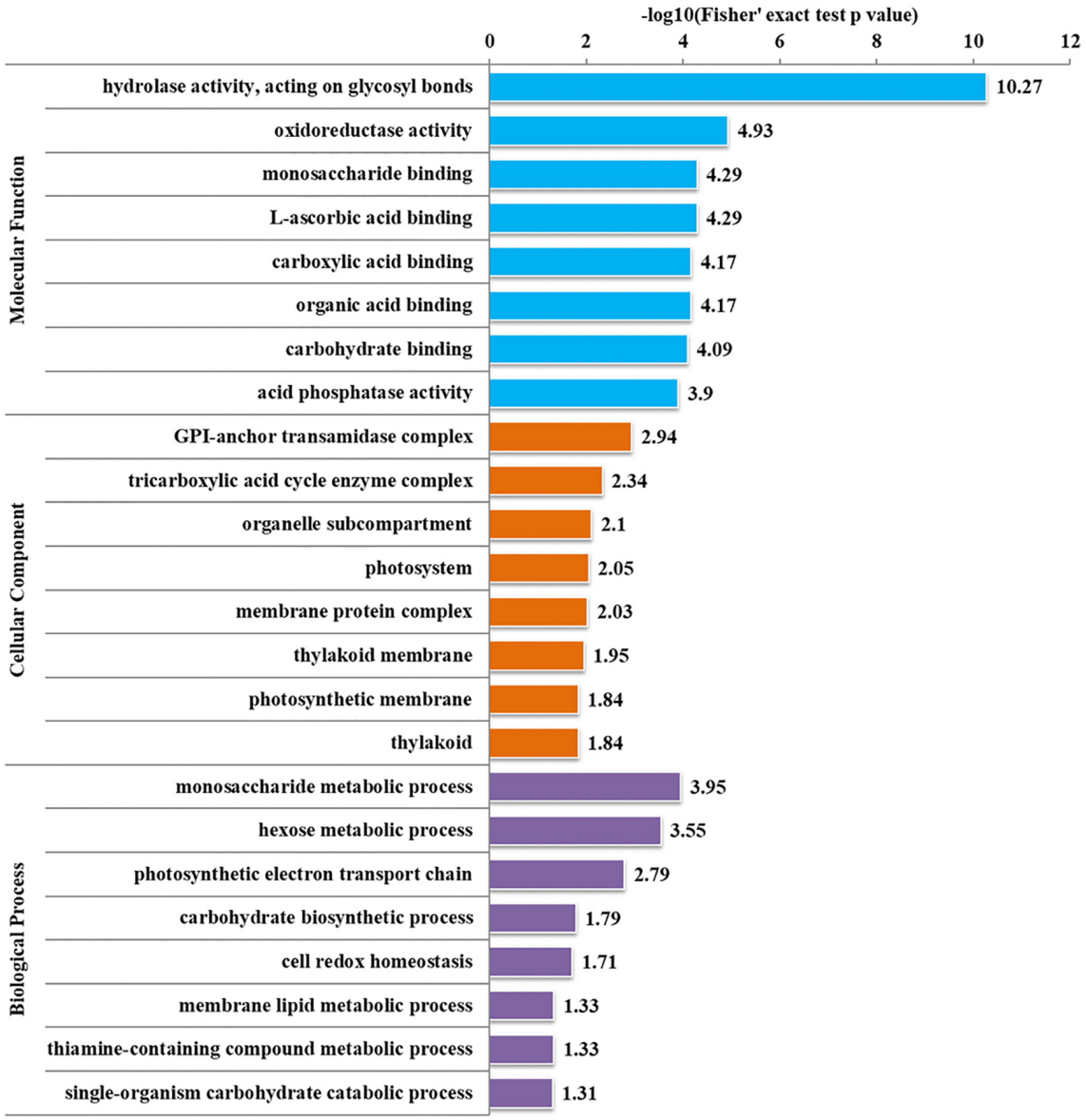


FIGURE 4 | Enrichment analysis of the N-glycoproteins based on Gene Ontology (GO) annotation. Gene ontology annotation based on the three main catteries: "Biological Process," "Cellular Component," and "Molecular Function." –log10 (Fisher's exact test *p*-value): *p*-value is adjusted by Fisher. The −log10 (Fisher' exact test *p* value) is higher, the enrichment is more significant.

Interaction Networks of N-Glycoproteins

N-linked glycosylation can significantly change the conformation of proteins and subsequently influence the interaction and function of proteins (Delgado-Vélez et al., 2021). Therefore, the analysis of N-glycoproteins interaction is important for the functional study of these N-glycoproteins. In this study, the PPI networks for all N-glycosylated proteins were established (**Figure 6**). A total of 74 N-glycoproteins were identified as nodes and connected as six interactional clusters (I–VI), as shown in **Figure 6**. The size of the node was the key parameter to evaluate the number of N-glycosylated modification sites. The biggest cluster was cluster I, containing 44 N-glycoproteins. Among them, glutamine synthase (GS, B7G6Q6) showed the largest size (five modification sites). Fumarate reductase flavoprotein (B7FVF5) showed the second-largest size (four modification sites). The cluster I contained several nodes involving in the biosynthesis of amino acids, such as glutamine synthase (GLNA, B7G6Q6), glutamine synthetase (GS, B7G5A1), and tryptophan synthase subunit beta (TrpB, B7FQI2). Glyceraldehyde-3-phosphate dehydrogenase, cytosolic class II aldolase, fructose-bisphosphate aldolase, and phosphoglycerate mutase were annotated to be involved in glycolytic pathway. Transketolase is part of the Calvin cycle. Furthermore, 11, 7, 5, 4, and 3 N-glycoproteins were observed in clusters II,

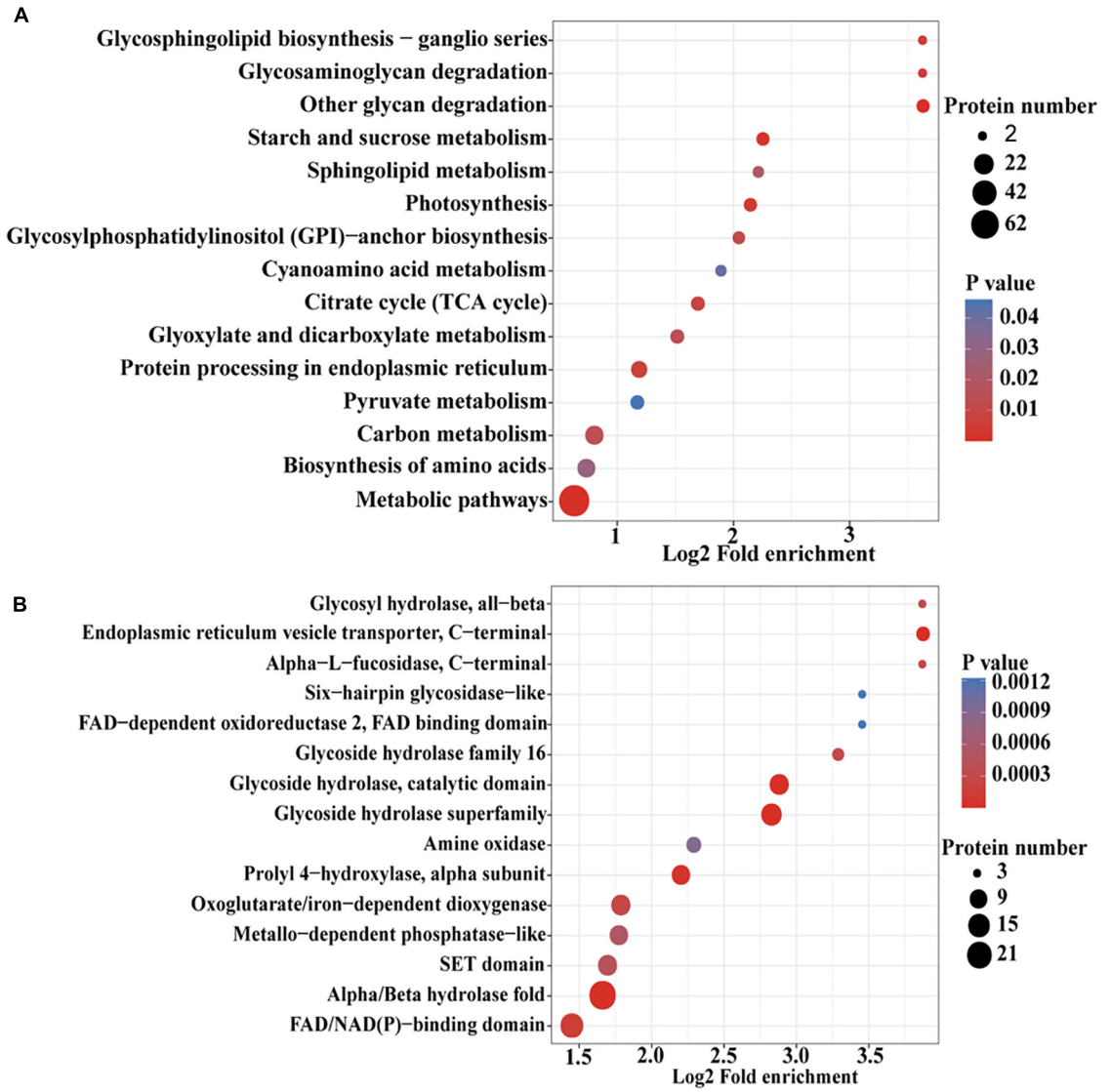


FIGURE 5 | An enrichment analysis of the N-glycoproteins based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and protein domains in *P. tricornutum*. **(A)** Enrichment of N-glycoproteins based on KEGG pathways. **(B)** Enrichment of N-glycoproteins based on protein domains.

III, IV, V, and VI, respectively. The cluster V harbored four N-glycoproteins, involving in starch and sucrose metabolism. They were glycoside hydrolase (ID: 56506), beta-glucosidases (ID: 26742 and ID: 50351), and exo-1,3-beta-glucosidase (ID: 49610), respectively (**Figure 6**).

N-Glycoproteins Involved in the Proposed N-Glycosylation Pathway

The pathway of protein N-glycosylation occurred in the ER and Golgi apparatus (**Figure 7**). In this study, 12 proteins involved in the N-glycosylation pathway are identified as N-glycoproteins, as shown in **Table 2**. Among them, ALG12, staurosporine and temperature-sensitive 3 (STT3) of oligosaccharyltransferases (OSTs) complex, binding protein (BiP), glucose regulating protein 94 (GRP94), protein disulfide isomerase (PDI), ER degradation-enhancing α-mannosidase-like protein (EDEM), Heat shock protein 90 (Hsp90), and core α-1,3-fucosyltransferase (FucT) N-glycoproteins contained one N-glycosylated modification site. While nucleotide exchange factor (NEF), Heat shock protein 70 G (HSP70G), Heat shock protein 70 A (HSP70A), and ubiquitin regulatory X (Ubx) N-glycoproteins harbored two N-glycosylated modification sites in *P. tricornutum*. ALG12, STT3 of OSTs complex, NEF, HSP70G, BiP, GRP94, and PDI are proposed to be involved in the ER pathway of protein N-glycosylation. However, FucT being putatively located to the Golgi apparatus, relating with the Golgi pathway of protein N-glycosylation. EDEM, Hsp90, HSP70A, and Ubx belonged to the ERAD pathway. Based on the prediction of PPI networks, it is found that eight identified N-glycoproteins

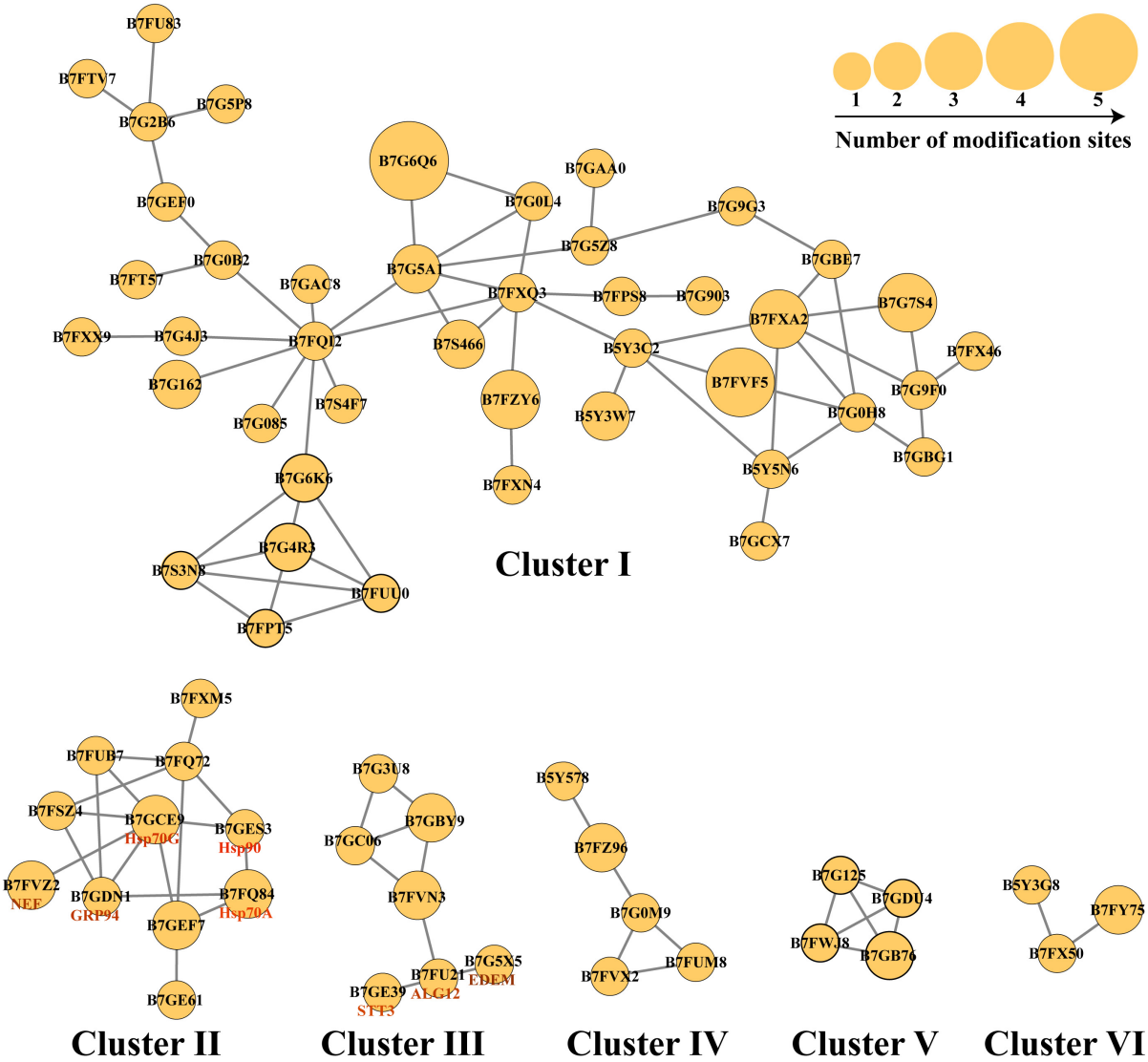


FIGURE 6 | The protein–protein interaction (PPI) networks. The PPI networks were divided into six interactional clusters, from cluster I to cluster VI. Each node is a N-glycoproteins. In total, 74 N-glycoproteins were identified from six clusters. B7FXM5 is mitochondrial chaperonin. B7FUB7 is endoplasmic reticulum (ER) luminal binding protein (BiP). B7FQ72 is mitochondrial chaperonin, Cpn60/Hsp60p. B7FSZ4 is a predicted protein. B7GEF7 is Heat shock protein Hsp90. B7GE61 is Heat shock protein Hsp70. B7G3U8 is a predicted protein. B7GBY9 is glycosylphosphatidylinositol (GPI)-anchor transamidase. B7GC06 is GPI transamidase complex, GPI16/PIG-T component, involved in GPI anchor biosynthesis. B7FVN3 is mannosyltransferase. The number represents the number of modified sites.

involved in protein N-glycosylation pathway are predicted to two interactional complexes (**Figure 6**). In the cluster II, NEF and Hsp70G, Hsp70G and Hsp90, Hsp90 and Hsp70A proteins might interact, respectively. GRP94 protein is proposed to interact with Hsp70G and Hsp70A proteins, but not with Hsp90. In the cluster III, STT3 and ALG12 might interact each other, ALG12 and EDEM proteins might interact each other.

N-Glycans Profiling

To acquire a global view of N-glycans in *P. tricornutum*, total proteins were extracted and digested by trypsin, and the digested peptides were desalted using homemade C18 SPE tips. The enriched N-glycopeptides *via* ZIC-HILIC SPE tips were subsequently isotopic labeled and analyzed by using C18-RPLC-MS/MS (HCD). Totally, 69 N-glycans corresponding to 59 N-glycoproteins were identified from this study. The detailed data are shown in **Supplementary Table 3**. All 69 N-glycans were divided into two classes, high mannose, and hybrid type N-glycans. One hybrid type N-glycan with a GlcNAc was identified as $GlcNAc_1Man_4GlcNAc_2$ (**Figure 8A**). Sixty-eight high mannose type N-glycans were identified, such as three $Man_5GlcNAc_2$ (4.35%, Man, mannose; GlcNAc, N-acetylglucosamine) (**Figure 8B**), three $Man_7GlcNAc_2$ (4.35%) (**Figure 8C**), 55 $Man_9GlcNAc_2$ (79.71%) (**Figure 8D**), and seven $Glc_1Man_9GlcNAc_2$ (10.14%, Glc, glucose) (**Figure 8E**).

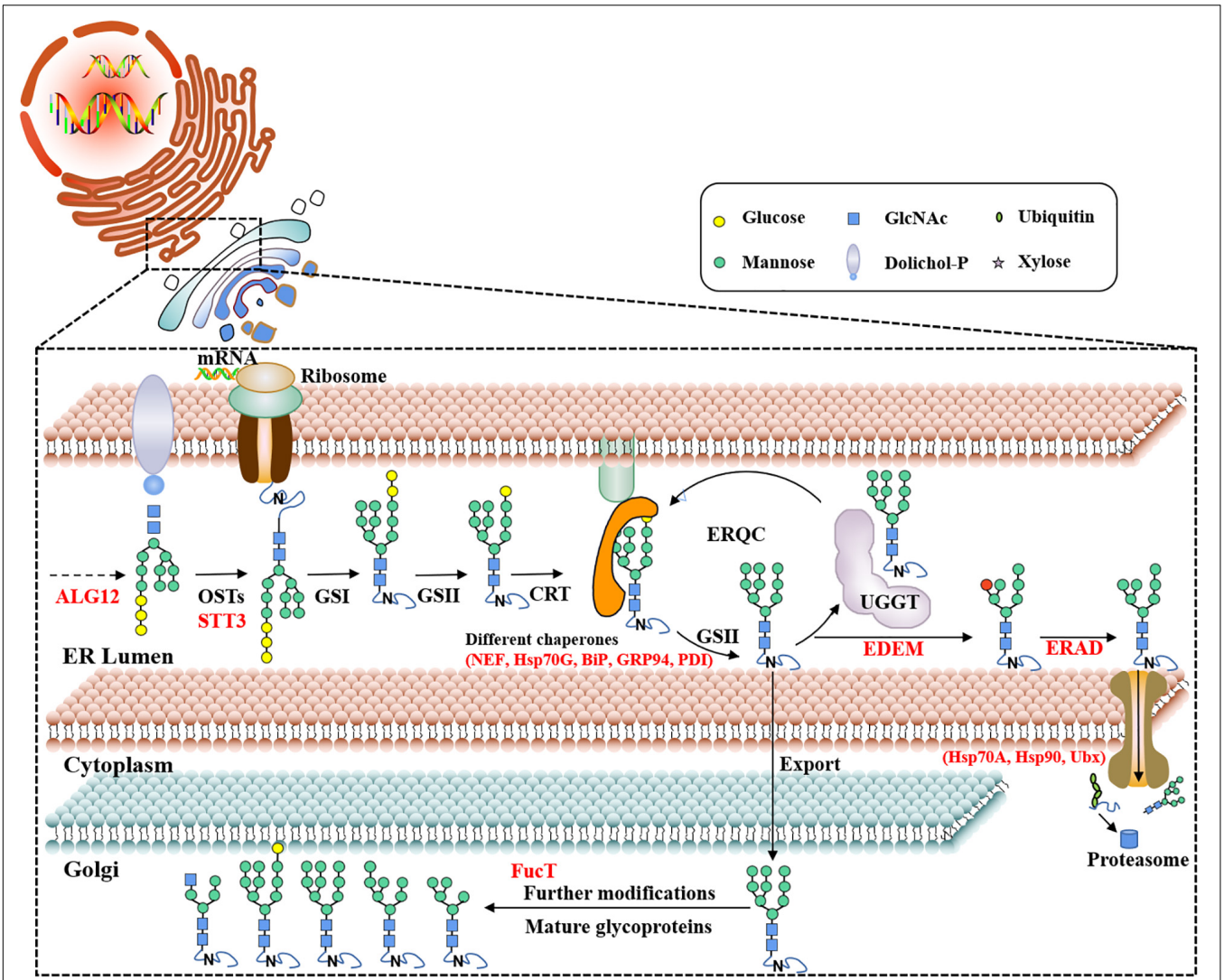


FIGURE 7 | Proposed protein N-glycosylation pathway and relevant N-glycoproteins in *P. tricornutum*. ALG12, asparagine-linked glycosyltransferase 12; OST, oligosaccharyltransferase complex; STT3, staurosporine and temperature-sensitive 3; UGGT, UDP-glucose: glycoprotein glucosyltransferase; CRT, calreticulin; GS I/II, glucosidase I/II; ERQC, ER quality control; EDEM, ER degradation-enhancing α-mannosidase-like protein; ERAD, ER-associated degradation; NEF, nucleotide exchange factor; Hsp70G, Heat shock protein 70 G; BiP, binding protein; GRP94, glucose regulating protein 94; PDI, protein disulfide isomerase; Hsp70A, Heat shock protein 70 A; Hsp90, Heat shock protein 90; Ubx, ubiquitin regulatory X; FucT, core α-1,3-fucosyltransferase.

DISCUSSION

We used an integrated N-glycoproteomic and N-glycomic approach to unravel the protein N-glycosylation pathway of *P. tricornutum* and shed light on N-glycans structure and N-glycoproteins. This approach has been used in limited studies so far. So far, the N-glycoproteomic analysis was performed in three green algae *C. reinhardtii* (Mathieu-rivet et al., 2013), *B. braunii* (Schulze et al., 2017), and diatom *T. oceanica* (Behnke et al., 2021). Compared with the identified 135 N-glycopeptides in *C. reinhardtii*, total 517 unique N-glycosylated peptides in three strains of *B. braunii* and 118 N-linked glycosylated peptides in *T. oceanica*, 863 different N-glycopeptides corresponding to 639 proteins were identified in this study in of *P. tricornutum*. The large difference on the number of the identified N-glycopeptides might be related with different methods. In *C. reinhardtii* and three *B. braunii* strains, the enrichments of N-glycopeptides were performed by using agarose-bound concanavalin A lectin, and the analysis methods of N-glycopeptides were PNGase F/^{18}O-method and in-source collision induced dissociation (IS-CID) methods (Mathieu-rivet et al., 2013) and LC-MS (Schulze et al., 2017), respectively. Furthermore, solid–phase extraction, hydrophilic-lipophilic balance (HLB) cartridge enrichment, LC-MS/MS analysis were carried out in the N-glycopeptide analysis of *T. oceanica* (Behnke et al., 2021). These enrichment methods are N-glycan-specific, especially the lectin enrichment, while HILIC microcolumn was considered in the current to get more enriched N-glycopeptides. But still, few known N-glycoproteins

TABLE 2 | The detailed information of N-glycoproteins during the protein N-glycosylation pathway.

Protein name	Protein ID	Amino acid position	Localization probability	Modified sequence
ALG12	44425	527	1	VSNSSTYTHMLSESK
OSTs (STT3)	55198	826	0.96	MWNLINSNAVEELK
NEF	44879	317	1	NTGEFAPWALEELTLGNESSIAR
NEF	44879	56	1	EWTLLRENDTVAAGMHVR
Hsp70G	55122	126	1	HYPVRPVYNETR
Hsp70G	55122	435	1	HINSDESMALGAAFAGANISTAFR
BiP	54246	567	1	NGLESYLYNLK
GRP94	16786	41	1	ELISNASDALDKFR
PDI	2808	169	1	GVLANVSK
EDEM	52346	116	1	VDRNVSVFETNIR
Hsp70A	54019	356	1	VQSMLSEFFNGKEPCK
Hsp70A	54019	63	0.99	SQAAMNAHNTVFDAK
Hsp90	55230	95	0.63	VDLVNNLGTIAK
Ubx	34061	59	1	GGGSGLAVQPNMDEGPDRD
Ubx	34061	43	1	AVFGLAENATAEDSGQSR
FucT	46109	554	1	EFGLGWNHTAQTIQPTHLPR

remained identified, such as tRNA synthetase (ID: 43097), the putative mismatch repair protein MutS (ID: 47730), and symbiont-specific ubiquitin fusion degradation protein 1 (ID: 49319) (Peschke et al., 2013), indicating that the analysis method of N-glycoproteins still needs to be improved.

In this study, we found that the canonical N-X-T motifs accounted for 57.9% of all motifs while in diatom *T. oceanica*, the identified peptides had 81% N-X-T motifs (Behnke et al., 2021), indicating that N-X-T motif is the main N-glycosylation site in diatom, and its percentage varies species to species. The presence of main N-glycoproteins in chloroplast and cytoplasm indicates the regulatory role of N-glycosylation in chloroplast and cytoplasm. It is already known that N-glycosylation modification could directly affect the subcellular localization of proteins, especially for proteins passing through the ER-Golgi network (Friso and Van Wijk, 2015). Therefore, it is proposed that the N-glycosylation plays an important role in the subcellular localization of these 639 N-glycoproteins. The subcellular localization of some identified N-glycoproteins had already been verified by experimental data in *P. tricornutum*. The first one was aminoacyl-tRNA synthetase (Glx, ID: 48863). Consistent with the prediction of subcellular localization in this study, N-glycosylated Glx was located to the plastid stroma of *P. tricornutum* (Peschke et al., 2013). It is certified that bulky glycoproteins could be transported into the complex plastid of *P. tricornutum*, and speculated that these proteins are N-glycosylated on the outermost plastid membrane and thereafter transported into the plastid (Peschke et al., 2013). Postsynaptic density protein 95, discs large, zonula occludens 1 (PDZ) domain containing glycoprotein-2 (PDZ-2, ID: 47562) was predicted to localize in plasma membrane in this study. This prediction is consistent with our previous experimental result (Liu et al., 2016). On the contrary, some predicted localizations of glycoproteins are not consistent with the experimental results. For example, the prediction showed that Omp85 is located to the vacuole, while our previous experimental result showed the third outermost membrane localization in *P. tricornutum* (Liu et al., 2016). This discrepancy indicated that the results from bioinformatic prediction need to be verified by the experiments.

Furthermore, N-glycoproteins were annotated and enriched *via* different database. The interaction networks of N-glycoproteins were predicted. These annotations, enrichments, and prediction of PPIs provide advantages for the researchers to study the functions of these N-glycoproteins in different biological processes and the effects of N-glycosylation on these N-glycoproteins (Wang et al., 2013). Here, taking glutamine synthase as an example, GS harbored five N-glycosylated sites in *P. tricornutum*. It is already known that GS was one of the most important enzymes for the nitrogen assimilation and absorption (Liu et al., 2019). GS catalyzed the assimilation of ammonium into glutamate and produced glutamine in the cytoplasm, mitochondria, and chloroplast (Konishi et al., 2016; Smith et al., 2019). The function of GS isoenzymes on the regulation of nitrogen metabolism was widely studied in different organisms, such as in microalgae *C. reinhardtii* (Cabello et al., 2020) and macroalga *G. lemaneiformis* (Liu et al., 2019). However, the N-glycosylation of GS in *P. tricornutum* is reported for the first time in this study. However, the effects of N-glycosylation on the biological functions of GS still remain unknown.

In addition to the proposed protein N-glycosylation pathway in green algae, *C. vulgaris* and *C. reinhardtii*, the pathway in *P. tricornutum* was first described *via* bioinformatic analysis (Baïet et al., 2011), showing that it was a GnTI-dependent process. Subsequently, the pathway was updated and was compared with *C. reinhardtii* and mammals. The data revealed a species specific nature of N-glycosylation pathway (Liu et al., 2021). Moreover, the genes involved in protein N-glycosylation pathway were also identified in *T. oceanica* (Behnke et al., 2021) and in other 17 microalgae (Liu et al., 2021) *via* bioinformatic methods. Although the protein N-glycosylation pathways in these microalgae are proposed, they need to be elucidated by experiments in the future. In this study, 12 N-glycoproteins involved in the protein N-glycosylation pathway were identified, such as ALG12, STT3, NEF, Hsp70G, BiP, GRP94, PDI, EDEM, Hsp70A, Hsp90, and Ubx in ER, and FucT in Golgi apparatus. It was indicated that the N-glycosylation modification of these proteins might be important for their functions during the protein N-glycosylation pathway. Two interaction networks of eight N-glycoproteins were predicted, suggesting that these N-glycoproteins might form two complexes and regulate the ERQC and ERAD in *P. tricornutum*. These results provide valuable data for studying function of genes (e.g., ALG12, STT3, and NEF) and the pathway of protein N-glycosylation. Until now, the functions of some genes involved in the protein N-glycosylation pathway have been reported. For examples, the heterologous expression of *P. tricornutum* GnTI restored the complex type N-glycans maturation in the Chinese hamster ovary (CHO) mutant, indicating its importance on the synthesis of complex type N-glycans (Baïet et al., 2011). Subsequently, it was observed that the GDP-fucose

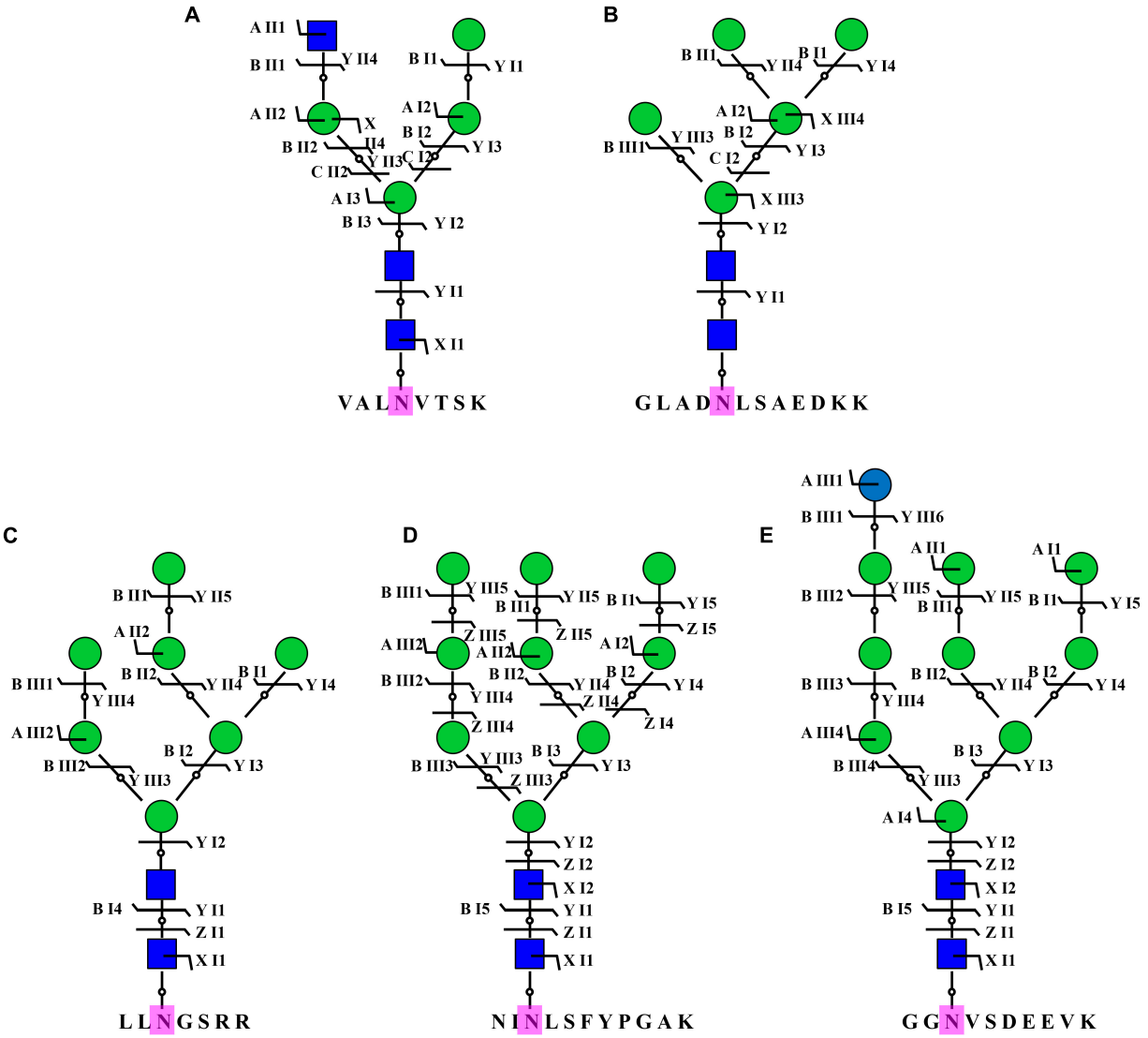


FIGURE 8 | Representative N-glycans identified in *P. tricornutum*. Five representative N-glycans were from five different N-glycoproteins, including B7FVG6 **(A)**, B7FUB7 **(B)**, B7FUL9 **(C)**, B7FUW2 **(D)**, and B7FTM0 **(E)**. Blue square is N-acetylglucosamine. Green circle is mannose. Blue circle is glucose. Purple N is asparagine amino acid for the N-glycosylation.

transporter and FucT1 were involved in the fucose modification of proteins in *P. tricornutum* (Zhang et al., 2019). In addition to *P. tricornutum*, the functions of genes, such as xylosyltransferase A (XylTA), xylosyltransferase B (XylTB), fucosyltransferase (FucT), and mannosidase 1A (Man1A) were studied during the protein N-glycosylation pathway of *C. reinhardtii* (Mathieu-Rivet et al., 2020). In *Porphyridium* sp., α-1,3-Glycosidase II was proposed to act in the ER (Levy-Ontman et al., 2015). The elucidations of genes function were important for revealing the protein N-glycosylation pathway in microalgae, especially for microalgae as cell factories to produce the humanized biopharmaceutical glycoproteins.

For the N-glycan structures, it was first reported that *P. tricornutum* proteins contained mainly high mannose type N-glycans, $Man_5GlcNAc_2$ to $Man_9GlcNAc_2$ (Baïet et al., 2011). The structure of the lipid linked oligosaccharide in *P. tricornutum* was $Glc_2Man_9GlcNAc_2$ (Lucas et al., 2018). These high mannose type N-glycans were further confirmed by IMS-MS and ESI-MS^n fragmentation patterns data (Dumontier et al., 2021). Subsequently, it was demonstrated that the proteins in *P. tricornutum* were N-glycosylated by single isomers of Man-5, Man-6, and Man-9 mannosides, and three isomers of Man-7 and Man-8 mannosides (Dumontier et al., 2021). Owing to the synthesis of canonical Man-5 and a GnTI-dependent pathway, it was speculated that *P. tricornutum* could initiate the processing of oligomannosides into complex type N-glycans (Baïet et al., 2011; Zhang et al., 2019). In this study, Man-5, Man-7, Man-9, and Man-10 mannose type N-glycans and a hybrid N-glycan

with a terminal GlcNAc residue were identified. However, different isomers were not found from these N-glycans. The more importance was that this study updated the N-glycans database of *P. tricornutum*, owing to the new reported N-glycans, such as $Glc_1Man_9GlcNAc_2$ and $GlcNAc_1Man_4GlcNAc_2$. Compared with the N-glycans reported in the previous studies, methylated mannose, linear $Man_5GlcNAc_2$, and complex N-glycans with xylose and/or fucose residues were not observed in this study because of two reasons. One reason might be owing to the loss caused by experimental operations. The other reason might be that PNGase F could not cleave complex N-glycans with fucose and/or xylose modified core structure (Song et al., 2013). Although PNGase A could cleave all types of asparagine binding N-glycans, the efficiency was very low (Song et al., 2013). Therefore, an appropriate identification method for N-glycans is very important.

Asparagine-linked glycosyltransferase 12, a N-glycoprotein, was the first enzyme in protein N-glycosylation pathway to be identified. ALG12 gene was predicted from the genome of most organisms, such as red microalga *Galdieria sulphuraria*, diatom *P. tricornutum*, green microalga *Ostreococcus lucimarinus*, and *Chlorella*, however, ALG12 was not annotated in the genomes of red microalgae *Cyanidischyzon merolae* and green microalgae *C. reinhardtii* (Levy-Ontman et al., 2014). It is reported that ALG12 is important for the formation of N-glycan structures, as the lack of ALG12 in *C. reinhardtii* the branched N-glycan is substituted by the linear oligomannosidic N-glycan (Mathieu-rivet et al., 2013). The mutation of ALG12 affected the transfer of mannose to N-glycan structure (Dempski and Imperiali, 2002). Additionally, it was indicated that ALG12 affected the N-glycan structure, but not the cell growth and viability (Kajiura et al., 2010). Therefore, it was speculated that the N-glycosylation of ALG12 might play an important role in the N-glycan structures in *P. tricornutum*. In conclusion, the exploration of the *P. tricornutum* N-glycoproteins, N-glycans, and the proposed N-glycosylation pathway as done in this study represents an important first step toward the design of genetically engineered driven remodeling of the alga to produce functional humanized biopharmaceutical N-glycoproteins for the clinical therapeutics.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

XL designed the experiments and supervised the project. XX, HoD, JC, MA, WW, WC, PL, HuD, and XL performed the experiments and the revised manuscript. XL and MA wrote the draft manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 41806168 and 41976125), Start-Up funding of Shantou University (grant number NTF18004), China Agriculture Research System (grant number CARS-50), Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (grant number GML2019ZD0606), and Department of Education of Guangdong Province (2018KCXTD012).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.779307/full#supplementary-material

REFERENCES

Baïet, B., Burel, C., Saint-Jean, B., Louvet, R., Menu-Bouaouiche, L., Kiefer-Meyer, M. C., et al. (2011). N-Glycans of Phaeodactylum tricornutum Diatom and Functional Characterization of Its N-Acetylglucosaminyltransferase I Enzyme. *J. Biol. Chem.* 286, 6152–6164. doi: 10.1074/jbc.M110.175711

Behnke, J., Cohen, A. M., and LaRoche, J. (2021). N-linked glycosylation enzymes in the diatom Thalassiosira oceanica exhibit a diel cycle in transcript abundance and favor for NXT-type sites. *Sci. Rep.* 11:3227. doi: 10.1038/s41598-021-82545-1

Blanco-Herrera, F., Moreno, A. A., Tapia, R., Reyes, F., Araya, M., D'Alessio, C., et al. (2015). The UDP-glucose: glycoprotein glucosyltransferase (UGGT), a key enzyme in ER quality control, plays a significant role in plant growth as well as biotic and abiotic stress in Arabidopsis thaliana. *BMC Plant Biol.* 15:127. doi: 10.1186/s12870-015-0525-2

Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, 344–354. doi: 10.1093/nar/gkaa977

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Cabello, P., Luque-Almagro, V. M., Roldán, M. D., and Moreno-Vivián, C. (2020). Chlamydomonas reinhardtii, an Algal Model in the Nitrogen Cycle. *Plants* 9:903. doi: 10.1016/B978-0-12-809633-8.20706-1

Cheng, A., Grant, C. E., Noble, W. S., and Bailey, T. L. (2019). MoMo: discovery of statistically significant post-translational modification motifs. *Bioinformatics* 35, 2774–2782. doi: 10.1093/bioinformatics/bty1058

Daboussi, F., Leduc, S., Maréchal, A., Dubois, G., Guyot, V., Perez-Michaut, C., et al. (2014). Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nat. Commun.* 5:3831. doi: 10.1038/ncomms4831

Delgado-Vélez, M., Quesada, O., Villalobos-Santos, J. C., Maldonado-Hernández, R., Asmar-Rovira, G., Stevens, R. C., et al. (2021). Pursuing high-resolution structures of nicotinic acetylcholine receptors: lessons learned from five decades. *Molecules* 26:5753. doi: 10.3390/molecules26195753

Dempski, R. E., and Imperiali, B. (2002). Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr. Opin. Chem. Biol.* 6, 844–850. doi: 10.1016/S1367-5931(02)00390-3

Du, H., Zheng, C., Aslam, M., Xie, X., Wang, W., Yang, Y., et al. (2021). Endoplasmic Reticulum-Mediated Protein Quality Control and Endoplasmic Reticulum-Associated Degradation Pathway Explain the Reduction of N-glycoprotein Level Under the Lead Stress. *Front. Plant Sci.* 11:598552. doi: 10.3389/fpls.2020.598552

Dumontier, R., Loutelier-Bourhis, C., Walet-Balieu, M. L., Burel, C., Mareck, A., Afonso, C., et al. (2021). Identification of N-glycan oligomannoside isomers in the diatom Phaeodactylum tricornutum. *Carbohydr. Polym.* 259:117660. doi: 10.1016/j.carbpol.2021.117660

Fang, H., Xiao, K., Li, Y., Yu, F., Liu, Y., Xue, B., et al. (2016). Intact Protein Quantitation Using Pseudoisobaric Dimethyl Labeling. *Anal. Chem.* 88, 7198–7205. doi: 10.1021/acs.analchem.6b01388

Friso, G., and Van Wijk, K. J. (2015). Posttranslational protein modifications in plant metabolism. *Plant Physiol.* 169, 1469–1487. doi: 10.1104/pp.15.01378

Hempel, F., Maurer, M., Brockmann, B., Mayer, C., Biedenkopf, N., Kelterbaum, A., et al. (2017). From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein. *Microb. Cell Fact.* 16:131. doi: 10.1186/s12934-017-0745-2

Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 158, 299–312. doi: 10.1104/pp.111.184333

Kajiura, H., Okamoto, T., Misaki, R., Matsuura, Y., and Fujiyama, K. (2012). Arabidopsis β1,2-xylosyltransferase: substrate specificity and participation in the plant-specific N-glycosylation pathway. *J. Biosci. Bioeng.* 113, 48–54. doi: 10.1016/j.jbiosc.2011.09.011

Kajiura, H., Seki, T., and Fujiyama, K. (2010). Arabidopsis thaliana ALG3 mutant synthesizes immature oligosaccharides in the ER and accumulates unique N-glycans. *Glycobiology* 20, 736–751. doi: 10.1093/glycob/cwq028

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. doi: 10.1002/pro.3711

Konishi, N., Ishiyama, K., Beier, M. P., Inoue, E., Kanno, K., Yamaya, T., et al. (2016). Contributions of two cytosolic glutamine synthetase isozymes to ammonium assimilation in Arabidopsis roots. *J. Exp. Bot.* 68, 613–625. doi: 10.1093/jxb/erw454

Levy-Ontman, O., Arad, S., Harvey, D. J., Parsons, T. B., Fairbanks, A., and Tekoah, Y. (2011). Unique N-glycan moieties of the 66-kDa cell wall glycoprotein from the red microalga Porphyridium sp. *J. Biol. Chem.* 286, 21340–21352. doi: 10.1074/jbc.M110.175042

Levy-Ontman, O., Fisher, M., Shotland, Y., Tekoah, Y., and Arad, S. M. (2015). Insight into glucosidase II from the red marine microalga Porphyridium sp. (Rhodophyta). *J. Phycol.* 51, 1075–1087. doi: 10.1111/jpy.12341

Levy-Ontman, O., Fisher, M., Shotland, Y., Weinstein, Y., Tekoah, Y., and Arad, S. M. (2014). Genes involved in the endoplasmic reticulum N-Glycosylation pathway of the red microalga Porphyridium sp.: a bioinformatic study. *Int. J. Mol. Sci.* 15, 2305–2326. doi: 10.3390/ijms15022305

Liu, M. Q., Zeng, W. F., Fang, P., Cao, W. Q., Liu, C., Yan, G. Q., et al. (2017). PGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun.* 8:438. doi: 10.1038/s41467-017-00535-2

Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., et al. (2016). Addressing various compartments of the diatom model organism Phaeodactylum tricornutum *via* sub-cellular marker proteins. *Algal Res.* 20, 249–257. doi: 10.1016/j.algal.2016.10.018

Liu, X., Huan, Z., Zhang, Q., Zhong, M., Chen, W., Aslam, M., et al. (2019). Glutamine Synthetase (GS): a Key Enzyme for Nitrogen Assimilation in The Macroalga Gracilariopsis lemaneiformis (Rhodophyta). *J. Phycol.* 55, 1059–1070. doi: 10.1111/jpy.12891

Liu, X., Xie, X., Du, H., Sanganyado, E., Wang, W., Aslam, M., et al. (2021). Bioinformatic analysis and genetic engineering approaches for recombinant biopharmaceutical glycoproteins production in microalgae. *Algal Res.* 55:102276. doi: 10.1016/j.algal.2021.102276

Lucas, P. L., Dumontier, R., Loutelier-Bourhis, C., Mareck, A., Afonso, C., Lerouge, P., et al. (2018). User-friendly extraction and multistage tandem mass spectrometry based analysis of lipid-linked oligosaccharides in microalgae. *Plant Methods* 14:107. doi: 10.1186/s13007-018-0374-8

Mathieu-Rivet, E., Mati-Baouche, N., Walet-Balieu, M. L., Lerouge, P., and Bardor, M. (2020). N- and O-Glycosylation Pathways in the Microalgae Polyphyletic Group. *Front. Plant Sci.* 11:609993. doi: 10.3389/fpls.2020.609993

Mathieu-rivet, E., Scholz, M., Arias, C., Dardelle, F., Schulze, S., Mauff, L., et al. (2013). Exploring the N -glycosylation Pathway in Chlamydomonas reinhardtii Unravels Novel Complex Structures. *Mol. Proteomics* 12, 3160–3183. doi: 10.1074/mcp.M113.028191

Mócsai, R., Blaukopf, M., Svehla, E., Kosma, P., and Altmann, F. (2020). The N-glycans of Chlorella sorokiniana and a related strain contain arabinose but have strikingly different structures. *Glycobiology* 30, 663–676. doi: 10.1093/glycob/cwaa012

Mócsai, R., Figl, R., Troschl, C., Strasser, R., Svehla, E., Windwarder, M., et al. (2019). N-glycans of the microalga Chlorella vulgaris are of the oligomannosidic type but highly methylated. *Sci. Rep.* 9:331. doi: 10.1038/s41598-018-36884-1

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, 182–185. doi: 10.1093/nar/gkm321

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

O'Neill, E. C., Kuhaudomlarp, S., Rejzek, M., Fangel, J. U., Alagesan, K., Kolarich, D., et al. (2017). Exploring the glycans of Euglena gracilis. *Biology* 6:45. doi: 10.3390/biology6040045

Peschke, M., Moog, D., Klingl, A., Maier, U. G., and Hempel, F. (2013). Evidence for glycoprotein transport into complex plastids. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10860–10865. doi: 10.1073/pnas.1301945110

Schulze, S., Oltmanns, A., Machnik, N., Liu, G., Xu, N., Jarmatz, N., et al. (2018). N-Glycoproteomic characterization of mannosidase and xylosyltransferase mutant strains of Chlamydomonas reinhardtii. *Plant Physiol.* 176, 1952–1964. doi: 10.1104/pp.17.01450

Schulze, S., Urzica, E., Reijnders, M. J. M. F., van de Geest, H., Warris, S., Bakker, L. V., et al. (2017). Identification of methylated GnTI-dependent N-glycans in Botryococcus brauni. *New Phytol.* 215, 1361–1369. doi: 10.1111/nph.14713

Shin, Y. J., Vavra, U., Veit, C., and Strasser, R. (2018). The glycan-dependent ERAD machinery degrades topologically diverse misfolded proteins. *Plant J.* 94, 246–259. doi: 10.1111/tpj.13851

Smith, S. R., Dupont, C. L., McCarthy, J. K., Broddrick, J. T., Oborník, M., Horák, A., et al. (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat. Commun.* 10:4552. doi: 10.1038/s41467-019-12407-y

Song, W., Mentink, R. A., Henquet, M. G. L., Cordewener, J. H. G., Van Dijk, A. D. J., Bosch, D., et al. (2013). N-glycan occupancy of Arabidopsis N-glycoproteins. *J. Proteomics* 93, 343–355. doi: 10.1016/j.jprot.2013.07.032

Strasser, R. (2016). Plant protein glycosylation. *Glycobiology* 26, 926–939. doi: 10.1093/glycob/cww023

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, 362–368. doi: 10.1093/nar/gkw937

Uitz, J., Claustre, H., Gentili, B., and Stramski, D. (2010). Phytoplankton class-specific primary production in the world's oceans: seasonal and interannual variability from satellite observations. *Global Biogeochem. Cycles* 24:GB3016. doi: 10.1029/2009GB003680

Vanier, G., Lucas, P. L., Loutelier-Bourhis, C., Vanier, J., Plasson, C., Walet-Balieu, M. L., et al. (2017). Heterologous expression of the N-acetylglucosaminyltransferase I dictates a reinvestigation of the N-glycosylation pathway in Chlamydomonas reinhardtii. *Sci. Rep.* 7:10156. doi: 10.1038/s41598-017-10698-z

Vanier, G., Stelter, S., Vanier, J., Hempel, F., Maier, U. G., Lerouge, P., et al. (2018). Alga-Made Anti-Hepatitis B Antibody Binds to Human Fcγ Receptors. *Biotechnol. J.* 13:1700496. doi: 10.1002/biot.201700496

Wang, H., Huang, H., Ding, C., and Nie, F. (2013). Predicting protein-protein interactions from multimodal biological data sources *via* nonnegative matrix tri-factorization. *J. Comput. Biol.* 20, 344–358. doi: 10.1089/cmb.2012.0273

Weyman, P. D., Beeri, K., Lefebvre, S. C., Rivera, J., Mccarthy, J. K., Heuberger, A. L., et al. (2015). Inactivation of Phaeodactylum tricornutum urease gene

using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol. J.* 13, 460–470. doi: 10.1111/pbi.12254

Xiao, J., Wang, J., Cheng, L., Gao, S., Li, S., Qiu, N., et al. (2020). A puzzle piece of protein N-glycosylation in chicken egg: n-glycoproteome of chicken egg vitelline membrane. *Int. J. Biol. Macromol.* 164, 3125–3132. doi: 10.1016/j.ijbiomac.2020.08.193

Xue, Y., Xie, J., Fang, P., Yao, J., Yan, G., Shen, H., et al. (2018). Study on behaviors and performances of universal: n-glycopeptide enrichment methods. *Analyst* 143, 1870–1880. doi: 10.1039/c7an02062g

Zhang, P., Burel, C., Plasson, C., Kiefer-Meyer, M. C., Ovide, C., Gügi, B., et al. (2019). Characterization of a GDP-fucose transporter and a fucosyltransferase involved in the fucosylation of glycoproteins in the diatom Phaeodactylum tricornutum. *Front. Plant Sci.* 10:610. doi: 10.3389/fpls.2019.00610

Zheng, C., Aslam, M., Liu, X., Du, H., Xie, X., Jia, H., et al. (2020). Impact of Pb on Chlamydomonas reinhardtii at Physiological and Transcriptional Levels. *Front. Microbiol.* 11:1443. doi: 10.3389/fmicb.2020.01443

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Xie, Du, Chen, Aslam, Wang, Chen, Li, Du and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 January 2022
doi: 10.3389/fpls.2021.760516

The Transition Toward Nitrogen Deprivation in Diatoms Requires Chloroplast Stand-By and Deep Metabolic Reshuffling

*Matteo Scarsini[1], Stanislas Thiriet-Rupert[1,2], Brigitte Veidl[1], Florence Mondeguer[3], Hanhua Hu[4], Justine Marchand[1] and Benoît Schoefs[1]**

[1] Metabolism, Bio-Engineering of Microalgal Molecules and Applications (MIMMA), Mer Molécules Santé, IUML—FR 3473 CNRS, Le Mans University, Le Mans, France, [2] Institut Pasteur, Genetics of Biofilms Laboratory, Paris, France, [3] Phycotoxins Laboratory, Institut Français de Recherche pour l'Exploitation de la Mer, Nantes, France, [4] Key Laboratory of Algal Biology, Chinese Academy of Sciences, Wuhan, China

Microalgae have adapted to face abiotic stresses by accumulating energy storage molecules such as lipids, which are also of interest to industries. Unfortunately, the impairment in cell division during the accumulation of these molecules constitutes a major bottleneck for the development of efficient microalgae-based biotechnology processes. To address the bottleneck, a multidisciplinary approach was used to study the mechanisms involved in the transition from nitrogen repletion to nitrogen starvation conditions in the marine diatom *Phaeodactylum tricornutum* that was cultured in a turbidostat. Combining data demonstrate that the different steps of nitrogen deficiency clustered together in a single state in which cells are in equilibrium with their environment. The switch between the nitrogen-replete and the nitrogen-deficient equilibrium is driven by intracellular nitrogen availability. The switch induces a major gene expression change, which is reflected in the reorientation of the carbon metabolism toward an energy storage mode while still operating as a metabolic flywheel. Although the photosynthetic activity is reduced, the chloroplast is kept in a stand-by mode allowing a fast resuming upon nitrogen repletion. Altogether, these results contribute to the understanding of the intricate response of diatoms under stress.

Keywords: carbon metabolism reorientation, stress, biotechnology, lipids, transcriptomics, turbidostat operated photobioreactor, pigments, photosynthesis

OPEN ACCESS

Edited by:
Jianping Yu,
National Renewable Energy Laboratory (DOE), United States

Reviewed by:
Changhong Yao,
Sichuan University, China
Inna Khozin-Goldberg,
Ben-Gurion University of the Negev, Israel

****Correspondence:***
Benoît Schoefs
benoit.schoefs@univ-lemans.fr

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *18 August 2021*
Accepted: *19 November 2021*
Published: *18 January 2022*

Citation:
Scarsini M, Thiriet-Rupert S, Veidl B, Mondeguer F, Hu H, Marchand J and Schoefs B (2022) The Transition Toward Nitrogen Deprivation in Diatoms Requires Chloroplast Stand-By and Deep Metabolic Reshuffling. Front. Plant Sci. 12:760516. doi: 10.3389/fpls.2021.760516

INTRODUCTION

Since the industrial revolution, the human population and cities size have risen. To sustain the continuous increase in the energy demand, fossil energies were massively used, becoming progressively limited. The accumulation of greenhouse gasses and pollutants in the atmosphere has serious environmental consequences to a point that alternative and more sustainable sources have to be developed (Moss et al., 2010; Levy and Patz, 2015). Being the only living organisms able to use inorganic carbon to produce organic compounds, photosynthetic organisms constitute an alternative choice. However, the large exploitation of plants for other purposes than food and feed production resulted in a counterproductive competition for

agricultural land use, rocking the inflation of food prices (Kazamia and Smith, 2014). Microalgae are another type of photosynthetic organisms that can be grown in open ponds or photobioreactors (*i.e.*, farmland areas, thus avoiding the land competition) with the aim to use them as cell factories in production platforms (Gordon et al., 2019). Among microalgae, diatoms represent one of the most abundant groups of marine eukaryotic organisms. Thanks to their unique ability to acclimate to the changing environmental conditions, they have colonized any kind of environment including lichens or mosses (Schoefs et al., 2020). Diatoms have always played a crucial role in the biosphere as they contribute to 20–40% of the oceanic biomass production and approximately 20% of the total carbon fixation (Benoiston et al., 2017; Falciatore et al., 2020). Their metabolic diversity and flexibility make them attractive organisms for a wide range of applications in diverse industrial and commercial fields, including biofuel production, nanotechnological applications, pigments and dietary lipids production (Mimouni et al., 2012).

Despite the tremendous progress in cultivation systems, the use of diatoms as cell factories remains in its infancy. Recent analyses pointed out the high cost of downstream processes and the relative lack of knowledge in diatom biology (Heydarizadeh et al., 2014; Vinayak et al., 2015). The actual view of the cell metabolic network is similar to a marshaling yard which distributes carbon atoms among the different biosynthetic pathways. In non-stressful conditions, the carbon atoms, our train carts, are mostly injected in the pathway producing carbohydrates, used as fuel for the mitochondria to get the energy needed for cell division. Stressful conditions trigger metabolic switches, changing the carbon flow toward the production of energy-dense compounds such as lipids (Sayanova et al., 2017). This metabolic reorientation appears as a default mechanism in diatoms accompanied by a reduction in the division rate (Heydarizadeh et al., 2019). In this study, the marine diatom *Phaeodactylum tricornutum* was selected as the model organism since it presents many advantages, including culture easiness and an annotated genome (Bowler et al., 2008; Butler et al., 2020; Falciatore et al., 2020). Furthermore, the taxon is genetically amenable (Apt et al., 1996; Karas et al., 2015; Serif et al., 2018; Sharma et al., 2018; Fabris et al., 2020; George et al., 2020; Hu and Pan, 2020). The effects of nitrogen starvation on *P. tricornutum* have been widely studied. Unfortunately, the lack of homogeneity in the experimental procedures and parameters does not allow to easily combine the different information to obtain a comprehensive view of the cellular responses from molecular to physiological levels (**Supplementary Table 1**). Thus, the dynamics and orchestration of the different cellular responses need further clarification. To significantly contribute to this field, a multidisciplinary strategy, including maintaining all parameters, but nitrogen availability, as stable as possible throughout the experiment was adopted: *P. tricornutum* was grown in a highly controlled photobioreactor operated in turbidostat mode, from an unlimited to depleted nitrogen conditions. Molecular, biochemical, and physiological parameters were followed along with the transition. This work shows that the nitrogen-depleted state contains cells in equilibrium with their environment, meaning that the cellular mechanisms no longer evolve qualitatively (change of the cell status), but only quantitatively (no change of the cell status). The qualitative variations, therefore, took place during a transition phase separating the initial equilibrium state (nitrogen repleted) and the final deficient state (N depleted). The transition is characterized by a carbon metabolism reorientation toward energy storage and a reduction of cellular activities, including plastidial ones that are kept on stand-by, waiting for the potential upcoming optimal growing conditions.

TABLE 1 | Summary of the cultivation parameters.

Parameter	Value
Culture volume	1 L
Cell concentration	$\sim$3.5 10^6 cells mL^{-1}
Carbon supply	0.24 mL min^{-1}
Temperature	21°C
Light	White LED panel, 175 μmol photons s^{-1} m^{-2}, light/dark cycle: 23h40/0h20

MATERIALS AND METHODS

The Photobioreactor

P. tricornutum (Pt4) (UTEX 646) was cultivated in an f/2 medium (Guillard and Ryther, 1962) in a photobioreactor (PBR) (FM150, Photon System Instruments, Czech Republic) operated in turbidostat mode (**Table 1**) using the company software (Version 0.7.14). The PBR was equipped with sensors for light transmission, chlorophyll (Chl) fluorescence, temperature-pH (Mettler Toledo InPro3253SG/120/PT1000), dissolved inorganic carbon (DIC) (Mettler Toledo InPro5000/12/120), and dissolved oxygen (Mettler Toledo InPro6800/12/120). The PBR setup is presented in **Supplementary Figure 1**. The culture was continuously supplied with 1.2 mL min^{-1} of a mixture of air and CO_2 (2,000 ppm CO_2 total). After 5 days of stable growing conditions, the PBR was fed with f/2 medium containing 0.15 mM $NaNO_3$. We will refer to the day after the medium was changed as the Day from the Reduction of Nitrogen Supply (DRNS). The pH, temperature, DIC, and O_2 were continuously monitored. The (Chl) fluorescence yield measurements were carried on in a light-adapted state and in a dark-adapted state of each light-dark cycle (**Table 1**) (for the definition of the fluorescence parameters, refer to **Supplementary Table 2**).

Estimation of the Growth Rate

The cultivation in turbidostat mode requires that the culture turbidity remains constant. Turbidity (OD735) was measured at 735 nm to avoid pigment contribution. To maintain the OD735 constant, a peristaltic pump (Photon Systems Instruments PP500) was actioned by the PBR central controller when OD735 increased over a threshold set at 0.151 to maintain a cell density of 4 10^6 $cells^{-1}$. The medium in excess was pushed in a container and collected as an overflow (14 in **Supplementary Figure 1**). The daily volume of overflow allows for the estimation of the

dilution rate. In fact, the time course of the cell density (x) is a function of the cell division rate (μ_s) and also of the dilution rate (D) (Equation 1) (Gresham and Hong, 2015).

$$\frac{dx}{dt} = \mu_s \cdot x - D \cdot x \quad (1)$$

In turbidostat mode, $\frac{dx}{dt}$ equals zero and the Equation 1 reduces to Equation 2.

$$\mu_s = D \quad (2)$$

The decrease in limiting nutrients induces a decrease in division rate that the automatic dilution compensates accordingly. Thus, the changes in the dilution rate reflect the division rate.

Cells were counted daily using a Neubauer hemacytometer according to Heydarizadeh et al. (2019).

Sample Harvesting

Samples were sterilely collected (**Supplementary Figure 1**). For carbohydrate, protein, lipid, and pigment extraction, cells were centrifuged (3,000 × *g*, 10 min, 4°C). Pellets were washed 3 times with a phosphate-buffered solution (PBS) before storage at −80°C. The samples for RNA extraction and metabolite extraction were collected on glass filters (Whatman® glass microfiber filters, Grade GF/A, Merck, Germany) under a gentle vacuum (50 mmHg), immediately frozen in liquid nitrogen, washed 3 times with a PBS solution, and stored at −80°C. Samples for carbon and nitrogen determination were harvested on fiberglass filters under gentle vacuum (50 mmHg), washed 3 times with a PBS solution, and dried at 70°C for 48 h. The sampling volumes and quantities are reported in **Supplementary Table 3**.

Elemental Analyses

The nitrate concentration in the medium was determined according to Carvalho et al. (1998). Briefly, the absorbance at 220 nm was measured on a centrifuged culture supernatant (Lambda-25, Perkin-Elmer, USA). A sodium nitrate ($NaNO_3$) aqueous solution was used for calibration.

Internal Nitrogen (N-int) and Carbon (C-int) pools were determined using the C7N elemental analyzer (Eager 300, Thermo Scientific, Massachusetts, USA) according to Heydarizadeh et al. (2019).

RNA Extraction and High Throughput Transcriptome Analysis

The total RNA was extracted from samples with the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Missouri, USA). The purified total RNA extracts were processed with an RNAstable® LD kit (Sigma-Aldrich). Paired-end next-generation sequencing has been performed by the Annoroad company (Corporate HQ, Beijing, China) with an Illumina NextSeq 550AR Sequencer. The sequencing results and quality are reported in **Supplementary Dataset 1**. The raw data were processed with the Trimmomatic tool (version 0.36, Bolger et al., 2014) to remove adapters. A total of 45.9 million high-quality clean reads (96.15% clean reads) were aligned with the Hisat alignment software (version 2.1.0, Kim et al., 2019) on the high-quality reference transcriptome (NCBI assembly accession number: GCF_000150955.2). Expression abundances were quantified using Htseq-count (version 0.11.2, (Anders et al., 2015) and differentially expressed genes (FDR ≤ 0.001; |Log2(FC)| ≥ 1) identified using the R package DESeq (Bioconductor version 3.11, Love et al., 2014). Functional annotations of the transcriptome were carried out using Gene Ontology (GO) terms and enriched with the bioinformatics tool Blast2Go and a manual cross-reference between published data and online public databases (https://protists.ensembl.org/; https://mycocosm.jgi.doe.gov/; https://www.ncbi.nlm.nih.gov/). Heatmaps were generated with an *ad hoc* Python script based on the seaborn repository (gplots version 0.11.0, https://seaborn.pydata.org/index.html). The scripts can be found in the following repository: https://github.com/IlBarbe/SimpleSmallPythonScipts.

Neutral Lipid Quantification

Neutral lipids were estimated by Nile red fluorescence according to Huang et al. (2019). Pellets were resuspended in dimethyl sulfoxide (DMSO) 25% and stained with 1.5 µL Nile red (Sigma-Aldrich, Saint-Quentin Fallavier, France) solution 1 mg mL^{-1} in DMSO (100%). The 298 K fluorescence at 580 nm was excited at 530 nm (slits 5 nm) (Perkin Elmer LS-55, Villebon sur Yvette, France). DMSO diluted glyceryl trioleate (Sigma-Aldrich T7140-500MG, purity ≥99%) was used for calibration (**Supplementary Figure 2**).

Protein Quantification

Protein extraction was performed by mechanical lysis with the FastPrep-24™ 5G Homogenizer (MP Biomed, France). Sample pellets were mixed with glass beads (Merck, USA, 425-600 µm, acid-washed) and 50 µL ultrapure water for mechanical lysis (4 m s^{-1} for 30 s). The sample was chilled in ice for 1 min. After a second lysis cycle, 950 µL of ultrapure water were added to the lysate and homogenized for 30 s. The lysate was centrifuged at 16,000 × *g* for 20 min at 4°C. Protein concentrations were determined using the Bradford method (Bradford, 1976) using a double-beam spectrophotometer (Perkin Elmer Lambda-25). Bovine serum albumin (Sigma-Aldrich, purity ≥ 98%) was used for protein quantification calibration (**Supplementary Figure 2**).

Carbohydrate Quantification

Carbohydrate quantification was performed according to a slightly modified Dubois' protocol (Dubois et al., 1956). Briefly, the extraction was performed by mechanical lysis with the FastPrep-24™ 5G Homogenizer (MP Biomed, France). Sample pellets were mixed with glass beads (Merck, 425–600 µm, acid-washed) and 1.5 mL of 0.05 M sulphuric acid-water solution. Extract-containing tubes were incubated at 60°C for 30 min, vortexing every 10 min, then chilled for 10 min on ice. The lysate was centrifuged at 16,000 × *g* for 20 min at 4°C. In glass tubes, 2.5 mL of H_2SO_4 (>95%) were added to 1 mL of extract supernatant with the subsequent addition of 250 µL of 3%, phenol-water solution (99%, Merck). The glass tubes incubated in a water bath at 95°C for 30 min and chilled on ice for 10 min.

Colorimetric quantification has been performed reading the absorbance at 485 nm using a double-beam spectrophotometer (Perkin Elmer Lambda-25). Glucose (Merck, USA, ≥ 99.5%) was used for total carbohydrate quantification calibration (**Supplementary Figure 2**).

Stored Energy Estimation

The energy stored in organic molecules (E_a) was obtained by converting the different energy storage fractions (*i.e.*, neutral lipids, proteins, and carbohydrates) into energetic equivalents using their combustion energy (*i.e.,* Lipids: 39,500 mJ mg^{-1}, Proteins: 24,000 mJ mg^{-1}, Carbohydrates: 17,500 mJ mg^{-1}) (Gnaiger, 1983; Aderemi et al., 2018).

The available energy is thus:

$$E_a(\text{mJ } 10^6\text{cells}^{-1}) = E_{\text{carbohydrates}} + E_{\text{lipids}} + E_{\text{proteins}} \quad (3)$$

where $E_{\text{carbohydrates}}$, E_{lipids}, and E_{proteins} were calculated by multiplying the respective combustion energy by the quantity of corresponding macromolecules *per* 10^6 cells.

Pigment Quantification

Pigment extraction was performed by mechanical lysis with the FastPrep-24™ 5G Homogenizer (MP Biomed, France) under green ambient light (Schoefs, 2004). Sample pellets were mixed with glass beads (Merck, 425-600 µm, acid-washed) and 100 µL of pure acetone (Merck, ≥ 99.5%) for mechanical lysis (4 ms^{-1} for 30 s). Samples were centrifuged at 16,000 × *g* for 5 min at 4°C and supernatants were collected in a new microtube. Lysis cycles were repeated until the pellets fade to white. Absorbance spectra of the pooled supernatants were recorded between 400 nm and 800 nm (Perkin Elmer Lambda-25). The Chl *a*, Chl c, and total carotenoids concentrations were calculated according to Heydarizadeh et al. (2017) using the following equations:

$$\text{Chl } a\ (\text{mg L}^{-1}) = \frac{[11,77\,(DO_{663nm} - DO_{750nm}) - 0.32\,(DO_{663nm} - DO_{750nm})] \times v}{l \times V} \quad (4)$$

$$\text{Chl c}\ (\text{mg L}^{-1}) = \frac{[26.27\,(DO_{630nm} - DO_{750nm}) - 3.52\,(DO_{663nm} - DO_{750nm})] \times v}{l \times V} \quad (5)$$

$$\text{Total carotenoids}\ (\text{mg L}^{-1}) = \frac{\left[1000\,(DO_{473nm} - DO_{750nm}) - 8.08 \times Chl\,a \times l - Chl\,b \times l\right] \times v}{183.4 \times l \times V} \quad (6)$$

where *v* is the pigment extraction volume, V is the sampling volume, and l is the optical path.

Metabolomic Analysis

Metabolomic analyses were performed on samples (*i.e.,* CTRL 4, 5, 6, 7, 8 DNRS, and 10/11 DNRS) grouped on single sample. Metabolite extractions were performed according to Courant et al. (2013). Briefly, samples were plunged into 5 mL of a boiling mixture (ethanol-water, 75:25 v/v, 90–95°C) for 1 min, vigorously shaken for 3 s, and re-incubated for 1 min. The extraction process was then quenched by plunging tubes for 3 min in ethanol water solution (75:25 v/v) at −80°C. After filter removal, the tubes were centrifuged (4,500 × *g*, 5 min, 20°C). The supernatants were filtered on precombusted glass wool and collected in glass vials. The organic solvent phase was evaporated using a rotavapor (Rotavapor R-205 Evaporation System, Buchi; heating bath B-490, Buchi, Vacuum pump PC500, Vacuubrand) and lyophilized (Lyophilizer Christ Alpha 1-2 Plus, Christ; Vacuum pump E2M5, Edwards). The lyophilized extract was then resuspended with 1 mL methanol water solution (20:80 v/v) and stored at −80°C prior to filtration (mesh 0.2 µm). Liquid chromatography-high resolution tandem mass spectrometry (LC-HRMS/MS) analyses were performed with an Agilent 1290 Infinity II LC system coupled to a high-resolution time-of-flight mass spectrometer (Q-Tof 6550 iFunnel, Agilent technologies, CA, USA) equipped with a Dual Jet Stream® electrospray ionization source (positive mode). Detailed information on analyses parameters is reported in **Supplementary Information 1**. The Agilent Mass Hunter Workstation software (version B.07) was used to process the raw MS data, including the extraction of molecular features (MFs), generation of molecular formulas, library searching, and database searching. The identification of the most relevant entities was supported by MS information and searches in the METLIN database (http://metlin.scripps.edu) and in the Dictionary of Marine Natural Products (DMNP) library (Blunt et al., 2008; Wolfender et al., 2015).

The metabolomic analyses resulted in the detection of 4,267 compounds ranging from 0.1 and 2.8 kDa in size. To reduce the number of compounds to elucidate, the change in compound detected quantity as compared to the change in transcript abundance during the time course of the experiment. As a preliminary step, only compounds and transcripts which significantly changed in abundance, as compared to those under controlled conditions, at least at a time point, were retained. After this first screening, each compound pattern has been compared with each gene expression pattern. To be as stringent as possible, only the combinations which displayed a significant direct correlation were included (|slope|=1, from 0.95 to 1.05; $p < 0.05$). The results showed 419 compound-gene combinations which included 115 compounds. The 115 compounds were kept for further analyses and elucidation.

Statistics

ANOVA was performed using the online freeware (https://acetabulum.dk/anova.html). Principal component analyses (PCA) were performed with a compiled python script exploiting pandas (version 1.2.1), numpy (version 1.19.5), matplotlib (version 3.3.3), and scikit-learn (version 0.24.1) packages. Propagation of uncertainties in data processing was carried out following the mathematical rules (Arras, 1998). Linear and nonlinear fitting analyses were performed with the free version of CurveExpert software available for download (Hyams, D. G., CurveExpert software, http://www.curveexpert.net, 2010).

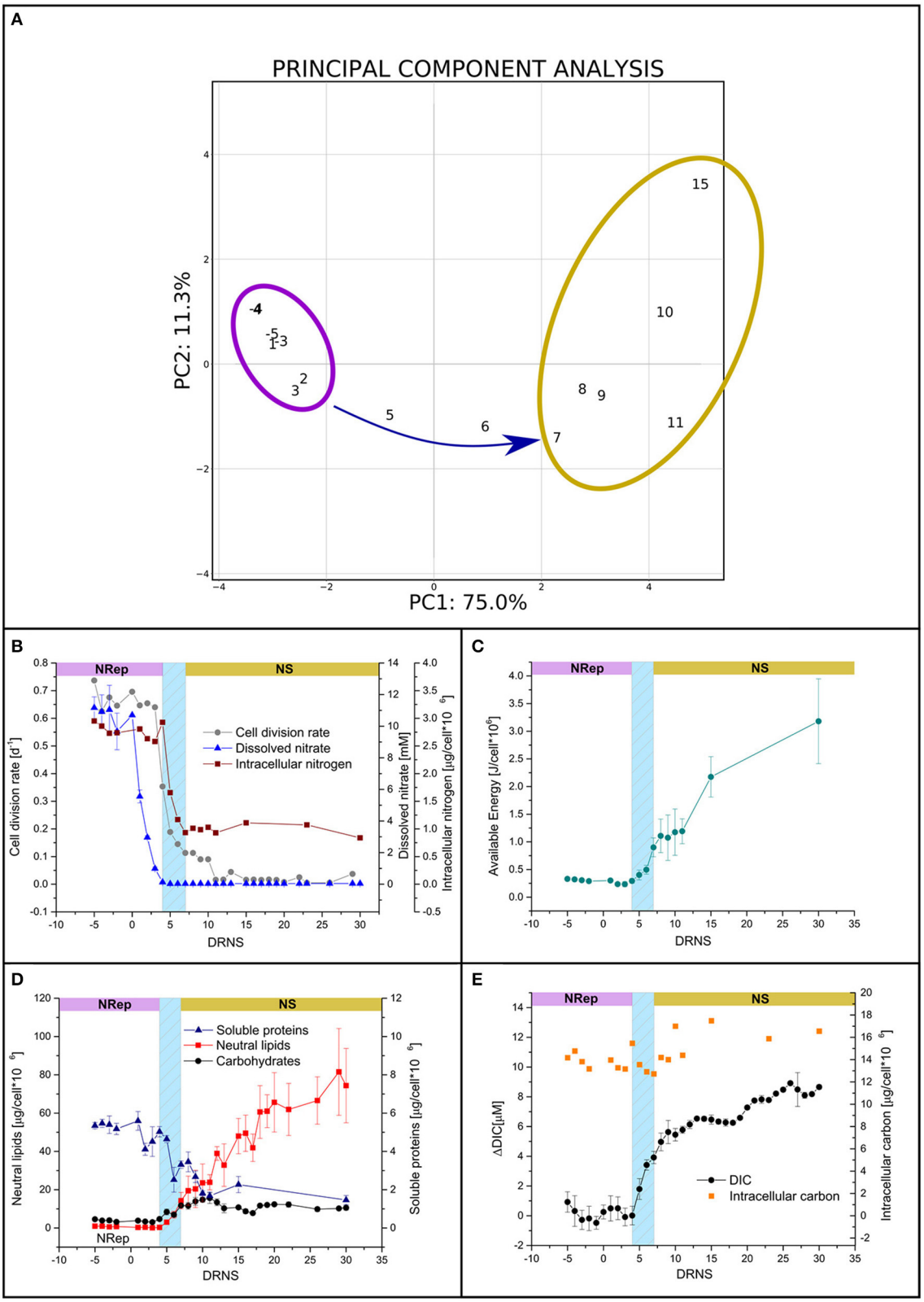


FIGURE 1 | Culture parameters in response to the transition from nitrogen-repleted (NRep) to nitrogen starvation (NS) conditions. **(A)** Principal component analysis (PCA) of the different physiological and biochemical traits. The first two components of the PCA explain 86,3% of the variance, 75% is explained by the first component allowing the outline of two groups: the NRep (Violet) and NS (Yellow) groups and the transition links them together. **(B)** Time-courses of cell density (grey circles), N-int (Red squares), and N-ext (blue triangles). **(C)** Cellular available biochemical energy during the time course of the experiment. Day 0 refers to DNRS. **(D)** Soluble proteins (blue triangles), carbohydrates (black circles), and neutral lipids (red squares) cellular quota. **(E)** DIC and intracellular carbon.

TABLE 2 | Summarizing table presenting the culture parameters during the NRep and NS phases.

Biochemical and Physiological trait	NRep	NS
Dissolved inorganic carbon availability variation [μM]	≃ 4	From 4 to 8
Extracellular inorganic nitrogen availability [mM of NO_3^-]	From 10 to 0.13	≃0.038
Neutral lipid cell quota [μg/10^6 cells]	≃ 0.5	From 14 to 80
Carbohydrate cell quota [μg/10^6 cells]	≃ 3.8	≃12
Soluble protein cell quota [μg/10^6 cells]	≃ 5.08	≃ 2.37
Average pigment cell quota [μg/10^6 cells]	Chl *a* = 0.18 Chl c = 0.024 Car = 19.12	Chl *a* = 0.06 Chl c = 0.016 Car = 7.08
Nitrogen cell quota [μg/10^6 cells]	≃ 2.78	≃ 0.99
Carbon cell quota [μg/10^6 cells]	≃ 13.99	≃ 14.86

RESULTS

P. tricornutum was cultivated in a PBR, operated in turbidostatic mode, and exposed to a 67-fold reduction in nitrogen availability. In order to identify the potential different steps of this transition, a PCA was performed with the biochemical and physiological traits (*i.e.,* DIC, neutral lipid, carbohydrate, protein, total pigment, N-int and C-int cellular quotas, Fv/Fm, qN, qP, and rETR). The first two components explained 86% of the sample variance (**Figure 1A**). Two groups can be defined: (i) a group containing all the time points from −2 to 4 DRNS that will be referred to as Nitrogen-Repleted (Nrep) condition, and (ii) a second group, from 7 to 15 DRNS that will be referred to as Nitrogen-Starved (NS) condition (**Table 2**). The two states are linked by a succession of transitory states through days 5 and 6, forming the transition phase. These two timepoints will be considered separately due to the importance of underlining the differences during the transition. The culture entered the transition phase, not at the onset of the DRNS but when the minimum N-ext availability is reached at 5 DNRS (0.03 mM NO_3^- corresponding to 0.007 μmol NO_3^- 10^{-6} cells) (**Figure 1B**). However, the transition between NRep and NS is rather led by N-int variations which is hereby considered as the key factor. Biochemical and physiological traits during the time course of the experiment were therefore plotted against the change in N-int. The plots, reported in **Supplementary Figure 3**, confirmed the information given by the performed PCA in which two groups are underlined with the transition in between. From the DRNS, N-ext decreased from 10 mM (NRep) to an average of 0.03 mM (NS) (**Figure 1B**; **Tables 1**, **2**). With a 4-day delay, the N-int dropped by 64%. Since culture entered the transition phase when minimum N-ext is reached and the NS phase when minimum N-int is reached (**Table 2**), N can be thus considered as the phase transition governing factor. The cell division rate decreased with timing and kinetic very close to the N-int (**Figure 1B**).

Effects of the Nitrogen Starvation on the Cellular Energy Content

Carbohydrates, lipids, and proteins are potential substrates for respiration and represent the three major macromolecule classes used to E_a. The kinetic of the E_a shows two increasing phases separated by a transient pause (ANOVA 95%, $p = 0.8$) (**Figure 1C**). The first increase depends on the transition phase and consists in the accumulation of both carbohydrates and neutral lipids (**Figure 1D**). A pause occurred at the onset of the NS phase with no significant E_a increase (ANOVA 95%, $p = 0.8$; **Figure 1C**). E_a accumulation resumed with the only energy accumulation in the form of neutral lipids (**Figure 1D**).

Soluble proteins cell quotas decreased gradually from the onset of the transition phase, negatively contributing to the energy storage (**Figure 1D**). In the present experiment, the major source of carbon as building blocks for these high-energy-containing molecules should have been DIC. However, DIC increased rapidly during the transition phase and has continued to increase in NS with a reduced rate (**Figure 1E**), while C-int displayed a small increase (9.3%), not coherent with the high lipid accumulation (**Figures 1C,D**). These observations suggest that the metabolic reorientation relies on carbon reallocation mechanisms rather than carbon fixation. To confirm the hypothesis of reduction in carbon fixation and bring evidence on the carbon reallocation theory, the photosynthetic activity and respiration activities were assessed.

Impact of the Nitrogen Starvation on the Photosynthetic and Respiratory Activities

In NRep, the cellular pigment quota remained stable and reduced to 37% of the maximum in NS conditions (**Table 2**). The reduction of the pigment cellular quota has direct effects on light energy harvesting and it can be a direct response to the changes in photosynthetic efficiency. Chlorophyll fluorescence was used as a proxy of photosynthetic efficiency (Roháček et al., 2008). Both the dark-adapted state (Φ_{P0}) and light-adapted state (Φ_{II}) decreased during the transition phase, reaching an average of 0.48 and 0.23 respectively (**Figure 2A**), indicating a reduced photochemical utilization of the light energy. The photochemical quenching (qP) that indicates the actual photochemical capacity of PSII in the light-adapted state decreased from 100% to 60–70%, suggesting a decrease in the capacity of reopening the PSII after excitation (**Figure 2B**). This generated a need to dissipate the excess of the absorbed energy as heat as revealed by the transient increase of the non-photochemical quenching (qN) from 5 to 15 DNRS (**Figure 2C**). A pattern similar to Φ_{P0} was observed for the relative electron transport rate (rETR) (**Figure 2D**).

The daily average dissolved oxygen rapidly decreased at the onset of the transition phase (**Supplementary Figure 4A**). The decrease in dissolved oxygen has a negative linear correlation with the increase in DIC (**Supplementary Figure 4B**), suggesting that the factor influencing the DIC affects also the dissolved oxygen. Since a negative linear correlation was observed between DIC and photosynthetic performance (**Supplementary Figures 5A,B**), the same correlation analysis

was performed between the dissolved oxygen and the same two photosynthesis parameters showing a positive relation in both cases (**Supplementary Figures 5C,D**). A decrease in dissolved oxygen is thus due to the decrease in its production by the reduced photosynthetic activity and its consumption by the respiratory activity.

Effect of a Small N Pulse on Cell Growth and Biochemical Composition

To get information on the capability of the cell to recover from NS, a nitrogen pulse was applied to the culture after 31 DRNS to reach a final nitrate concentration of 0.35 mM. **Supplementary Figure 6** shows the fast response of the cells to the sudden increase in N-ext. Within 24 h, cells remobilized the energy stored in lipids to transiently synthesized proteins and pigments, resuming cell division after a lag phase of at least 3 days as well.

Transcriptomic and Metabolomic Analyses

On the mindset of investigating the effects and mechanisms underlying the NRep-NS transition, transcriptomic analyses were performed on 11 time points. To observe the changes induced by the reduction of nitrogen supply (RNS), the changes in gene transcript availability were expressed compared with the control NRep. Overall, 2,560 genes were differentially expressed in at least one DNRS (**Supplementary Dataset 1**). The PCA based on transcriptomic data (**Figure 3A**) exhibits a pattern very similar to the PCA based on the biochemical and physiological data (**Figure 1A**). However, they differ by the positioning of 15 DNRS that is hereby detached from the NS phase group, suggesting that important changes in gene expression occurred at this time point. For the sake of simplicity in the figure representation, five phases will be considered: (i) NRep, (ii-iii) 5 and 6 DRNS, (iv) NS, and (v) 15 DRNS. In the case of NRep and NS, the average transcript levels across the corresponding DNRS were calculated for each gene for heatmaps creation (**Figures 4–6**). To strengthen the information given by the transcriptomic analysis, metabolic profiling analysis was performed through LC-MS on samples corresponding to 4, 5, 6, 8, 10, and 11 DNRS, which all allowed the detection of 4,267 compounds (p-value ≤ 0.05) (**Supplementary Dataset 2**). Individual compound cell quota change kinetics showed that 2,867 compounds displayed a significant variation ($|\log2FC| \geq 2$) compared to controlled conditions (before DRNS) for at least one time point. A PCA analysis was performed taking into account these 2,867 compounds (**Figure 3B**), thus, resulting in a similar pattern as observed on a transcriptomic level. A little shift was observed corresponding to the translation of the transcript changes at the metabolic level, and a great change observed was also observed on days 10 and 11.

Nitrogen Uptake, Scavenging, and Assimilation

At the onset of the transition phase, the transcript availability of genes encoding for plasma membrane nitrate and ammonium transporters has increased. The upregulation of the urea transporters (UT) encoding genes was observed when cells entered the NS phase (**Figure 4**). Similarly, an increase in the transcript levels is also observed in two of the three genes encoding for the vacuolar nitrate transporters. On the contrary, the expression level of the gene encoding the plastidial nitrite transporter (NAR1), nitrate reductase (NR), and the plastidial nitrite reductase (NiR) displayed a peculiar expression pattern, with a decrease during the transition phase and overexpression in NS.

Genes involved in the mitochondria ammonia fixation were mostly upregulated from the onset of the transition phase. A strong upregulation of the gene encoding the cytoplasmic glutamate dehydrogenase (GDH-B: J13951) was observed as soon as the NS state was reached. An increase in transcript levels was observed for the mitochondrial glutamine oxoglutarate aminotransferase (GOGAT) but not for the plastidial GOGAT encoding gene (**Figure 4**).

External amino acids could serve as an alternative nitrogen source (Flynn and Butler, 1986). In this concern, the periplasmic L-amino acid oxidase (PLAAOx) catalyzes the cleavage of the ammonium group from free amino acids (Thiriet-Rupert et al., 2018). The transcript levels of the PLAAOx encoding gene greatly increased from the onset of the transition phase and remained upregulated during NS. However, a significant increase in PLAAOx transcripts starting from 4 DRNS suggested a potential role in cell-sensing of N level. Thiriet-Rupert et al. (2018) reported that the PLAAOx encoding gene is coexpressed with the gene encoding the coccolith scale associated protein (CSAP) during NS in *Tisochrisis lutea*, which was also observed in *P. tricornutum* ($R = 0.9963$, $R^2 = 0.9926$).

An increase in transcript levels was observed for genes encoding proteins involved in nitrogen recycling such as deaminases, acetamide, and formamidase (Alipanah et al., 2015). The genes encoding for transporters of different nitrogen-containing molecules, such as polyamines and small polypeptides, were also upregulated in NS (**Supplementary Dataset 1**).

Amino Acid Metabolism

The total cell protein quota started to decrease after the RNS, reflecting both an increase in protein degradation and a decrease in protein synthesis. Upregulation of the genes encoding ubiquitin ligases, the enzymes that initiate the protein degradation and two 26S proteasome regulatory subunits (Subunit N3, EG02475, isoform 2 of the subunit 5, and EG0162) was observed in NS. Moreover, an increase in transcript levels encoding components of the ubiquitin-mediated protein degradation pathway was observed (**Supplementary Dataset 1**). Coherently, a decrease in transcript abundance for genes encoding aminoacyl-tRNA synthetases, ribosomal proteins, while translation and elongation factors were also observed. The overall transcript availability of genes involved in the biosynthesis of amino acids was reduced in NS conditions (**Supplementary Dataset 1**), with the sole exception of the serine and glycine, glutamine and glutamate biosynthetic pathway involved in photorespiration and central metabolism (**Supplementary Figures 8**, **9**).

Branched-Chain amino acids (BCAA) have been reported to be involved in the accumulation of triacylglycerols (TAG)

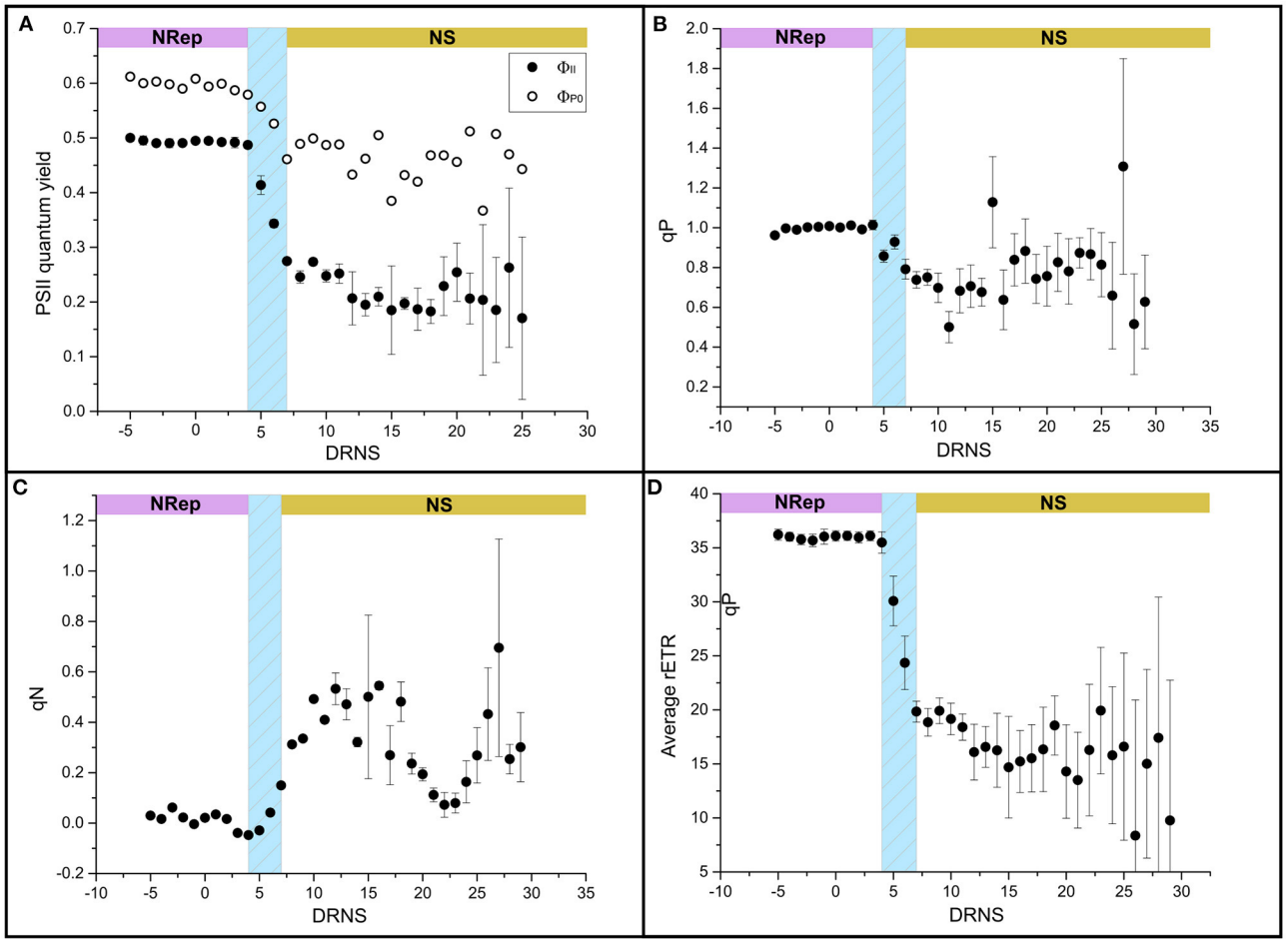


FIGURE 2 | Impact of NS on the photosynthetic performance. **(A)** Light-adapted state (ΦII) (closed symbol) and dark-adapted state (ΦP0) (open symbol), **(B)** photochemical quenching (qN), **(C)** photochemical quenching (qP), and **(D)** relative electron transport rate (rETR).

(Ge et al., 2014; Liang et al., 2019; Pan et al., 2020). Most of the genes involved in the BCAA degradation were upregulated (**Supplementary Figure 10**). In particular, the transcript abundance of the gene encoding the dihydrolipoyllysine-residue (2-methylpropanoyl) transferase (DHLTA, J54219), increased over 6 folds, remaining upregulated all throughout the transition and NS phases (D15 included) (**Supplementary Figure 10**). Transcript abundance of two genes encoding propionyl-CoA carboxylase carboxylases [PCC1 (J51245) and PCC2 (J45886)] started to be upregulated at the onset of the transition phase peaking during the NS. Furthermore, an increase in the transcript level was observed in the gene encoding the methylcrotonyl-CoA carboxylase (MCC2) (shikimate pathway) (**Supplementary Figure 10**). Pan et al. (2017) reported MCC2 as a key enzyme for which the silencing of the encoding gene (J19329) caused a decrease of carbon flux toward the TCA cycle, resulting in a decrease in TAG accumulation and nitrogen utilization. A change was also observed for the 3-hydroxyisobutyryl-CoA hydrolase (HIBCH) transcript abundance, which significantly increases in NS (**Supplementary Figure 10**).

Aside from proteins, other molecules such as polyamines require N for their synthesis. Polyamines are playing an essential regulatory function and are synthesized in all organisms including diatoms (Lin and Lin, 2019). The overall transcript abundance of the genes involved in the synthesis of the polyamines was decreased from the onset of the transition to the establishment of the NS phase (**Supplementary Dataset 1**).

Lipid Metabolism

Figure 5 presents the transcript abundance changes for the genes involved in lipid metabolism including *de novo* FA biosynthesis and the acetyl-CoA independent pathway.

De novo Biosynthesis

De novo FA biosynthesis depends directly on the plastid-residing acetyl-CoA pool, which is most likely the primary substrate for this pathway in diatoms (Zulu et al., 2018). *De novo* FA biosynthesis begins with the condensation of acetyl-CoA and malonyl-CoA, which are gradually added to the forming acyl molecule through the FA synthesis (FAS) cycle resulting in the acyl-CoA formation (Wang et al., 2018; Zulu et al., 2018).

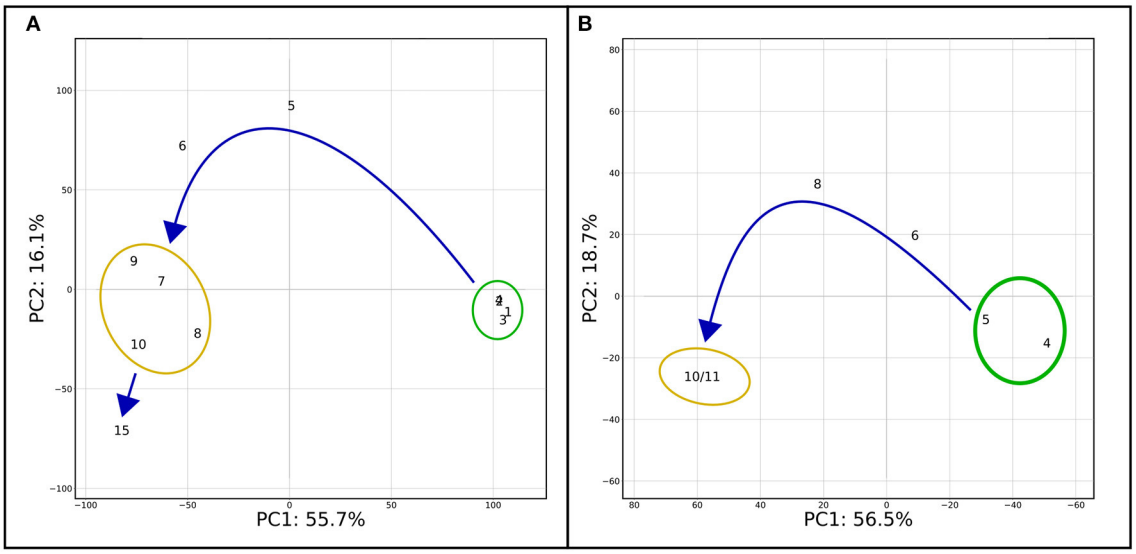


FIGURE 3 | PCA on transcriptomic and metabolomic results. **(A)** PCA was performed on the significantly differentially expressed genes (|log2FC| ≥ 1). The first two components of the PCA explained 71.8% of the variance. **(B)** PCA performed on metabolites metabolite abundance compared to the control (before DRNS) during the time course of the experiment. Only metabolites which change is significant (|log2FC| ≥ 2) in at least one of the time points were considered for the analysis. As for **Figure 1A**, two major groups are delimited: NRep (green) and NS (yellow) with a transition phase in between. Differently from **Figure 1A**, day 15 has been considered as separate point and not together with the NS group.

Then, the acyl groups are sequentially attached to the glycerol molecule, generated by the activity of the phosphoglycerol kinase (PGK) to form both storage and structural lipids (**Figure 5**). The whole genes encoding for components of the *de novo* FA biosynthesis were strongly downregulated (**Figure 5**). Moreover, the transcript abundance of most of the genes encoding enzymes, that involved in the acylglycerols formation, were reduced all along with the transition phases and the NS phase. In particular, all the five genes encoding lysophosphatidic acid acyltransferase (LPAT), responsible for the intermediate conversion of lysophosphatidic acid to phosphatidic acid, were significantly downregulated. The downregulation of the LPAT encoding gene under NS conditions has been observed by Alipanah et al. (2015), in opposition with what was observed previously by Valenzuela et al. (2012). The final conversion of DAG to TAG is carried out by the action of diacylglycerol *O*-acyltransferases (DGAT) encoded by six paralogous, four of which are differentially regulated. Two are slightly upregulated (DGAT1 and DGAT2D) while the other two are downregulated (DGAT2A and DGAT2C) in NS. The upregulation of DGAT1 and DGAT2D under NS starvation has been also observed by Alipanah et al. (2015). On the other hand, Remmers et al. (2018) reported no significant upregulation of the six DGAT encoding genes suggesting that the activity of these enzymes was not a limiting factor during steady-state growth under N-limitation conditions.

Acetyl-CoA Independent Biosynthesis

Besides the *de novo* FA biosynthetic pathway, a second pathway for FA synthesis is present in diatoms: the acetyl-CoA independent. The latter does not require the presence of an acetyl-CoA pool for the synthesis of acyl-CoA but recycles it from pre-existing polar lipids (Cagliari et al., 2011) (**Figure 5**). Membrane lipids including phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylinositol (PI), and phosphatidylglycerol (PG) are hydrolyzed by phospholipases. Phospholipases C (PLC) and phospholipases D (PLD) are responsible for the release of DAG from PI/PG and PC/PE, respectively (Alipanah et al., 2018). While the negligible change in transcript abundance was observed for the genes encoding PLC, a strong upregulation was observed for genes encoding the PLD. Five out of 6 PLD encoding genes were strongly upregulated from the onset of the transition throughout NS (**Figure 5**). The released DAG can serve as a substrate for TAG formation through the activity of DGAT and PDAT. The latter then catalyzes the transacylation of an acyl group in position *sn*-2 of the PC to position *sn*-3 of the DAG forming TAG and LPC. The transcript abundance of the single PDAT gene in *P. tricornutum* started to increase at the onset of the transition phase, reaching a maximum in NS. Another acyl group recycling pathway is the Land's cycle (Lands, 1960; Zulu et al., 2018) (**Figure 5**), which involves the hydrolysis of PC and PE in Lyso-PC (LPC) and Lyso-PE (LPE), respectively, by the activity of the phospholipases A (PLAs). The second reaction that closes the cycle is catalyzed by the activity of the acyl-CoA:lysophosphatidylcholine (LPC) acyltransferase (LPCAT), which leads to the reformation of PC and PE. While one single LPCAT has been identified in *P. tricornutum* (Bowler et al., 2008), two PLA transcripts have been identified encoding for two different PLAs: PLA1, which cleaved an acyl group in position *sn*-1 from a PC releasing the acyl-CoA and LPC, whereas PLA2 cleaved the acyl group in

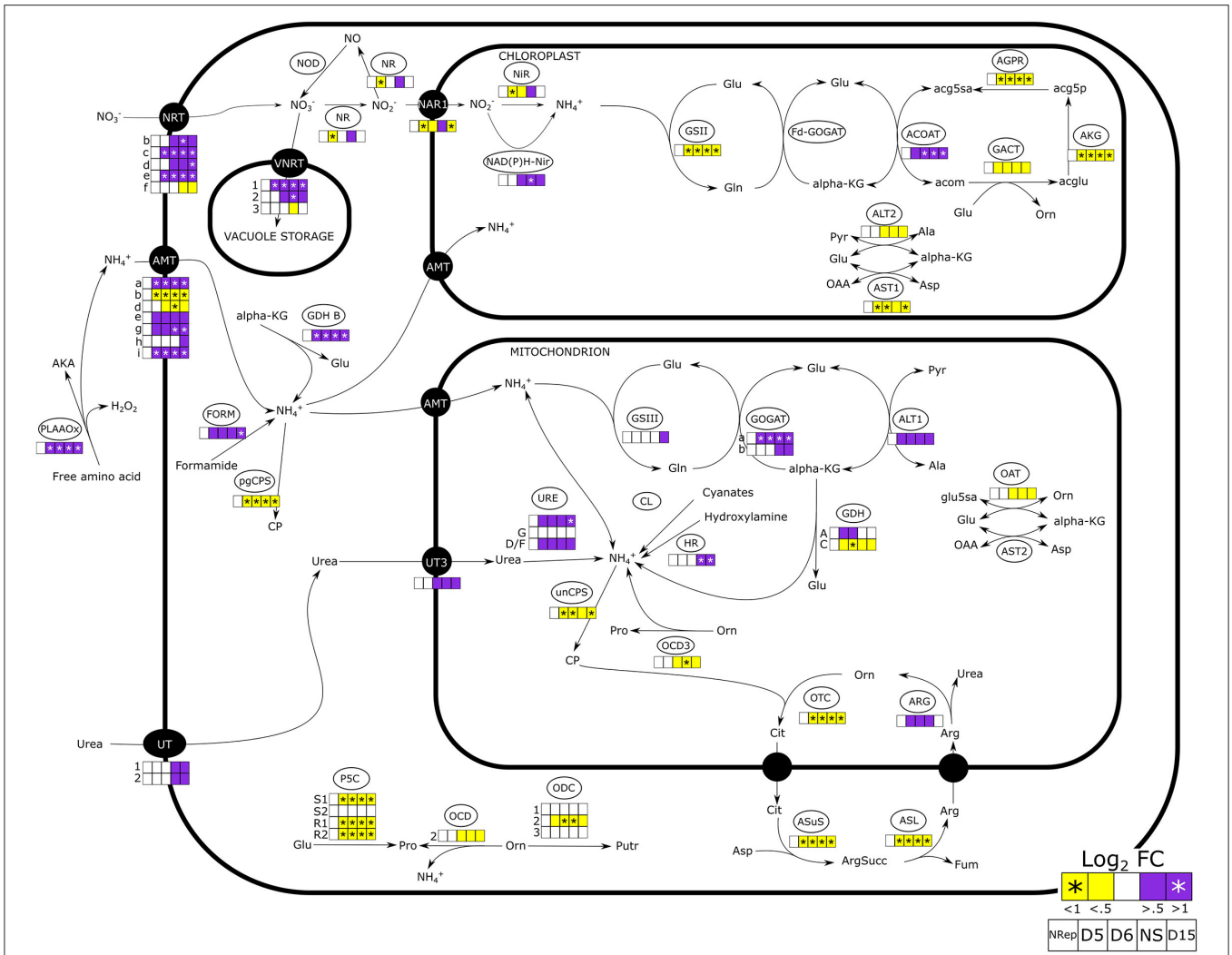


FIGURE 4 | Nitrogen uptake and assimilation pathways in *Phaeodactylum tricornutum*. The heatmaps report the change in gene expression in the 5 different groups defined through the PCA analyses (Nrep, D5, D6, NS, and D15). Daily variations of gene expression are reported in **Supplementary Figure 7**. NRT, nitrate transporter; VNRT, vacuolar nitrate transporter; NR, nitrate reductase; NAR1, nitrite transporter; Nir, nitrite reductase; NAD(P)H-Nir, nitrite reductase NAD(P)H dependent; GSII/III, glutamate-ammonia ligase; GOGAT, glutamate synthase; ACOAT, acetylornithine aminotransferase; AGPR, N-acetyl-gamma-glutamyl-phosphate reductase; GACT, glutamate N-acetyltransferase; AKG, acetylglutamate kinase; ALT, L-alanine transaminase; AST, aspartate aminotransferase; PLAAOx, periplasmic L-amino acid oxidase; FORM, formamidase; GDH, glutamate dehydrogenase; CP, carbamoyl phosphate; pgCPS, carbamoyl phosphate synthetase; unCPS, carbamoyl phosphate synthase mitochondrial, UT, urea transporter; URE, urease; CL, cyanate lyase, HR, Hydroxylamine reductase; GDH, glutamate dehydrogenase; OCD, ornithine cyclodeaminase; ODC, ornithine decarboxylase; P5CS, pyrroline-5-carboxylate synthase; P5CR, pyrroline-5-carboxylate reductase; OTC, ornithine transcarbamylase; ARG, arginase; ASL, argininosuccinate lyase; ASuS, argininosuccinate hydratase; AKA, α-keto acid.

position *sn*-2. Transcript abundance of genes encoding PLA1a and PLA2 significantly increased at the end of the transition phase (7 DRNS) and remained upregulated throughout the NS. In line with what was reported by Abida et al. (2015), a decrease in transcripts encoding MGDG synthase (MGD) and an increase in DGDG synthase (DGD) was observed, suggesting a reduction in the MGDG synthesis and its conversion to DGDG for further degradation. Decrease of membrane lipids, especially plastidial membrane lipids such as MGDG and DGDG, was also observed from a metabolomics point of view (**Supplementary Dataset 2**). Moreover, two putative DAG forms were shown to decrease in the NS phase and are in line with the hypothesis of the remodeling of lipids toward energy storage ones (**Supplementary Dataset 2**).

Lipid Storage

Due to their non-polar nature, neutral lipid accumulation requires proper storage in hydrophobic subcellular structures called lipid droplets (LD), which are characterized by specific LD proteins (LDP) (Goold et al., 2015; Lupette et al., 2019; Leyland et al., 2020). Two different LDP are present in *P. tricornutum*, StLDP (Stramenopiles LDP) and in PtLDP1 (*P.*

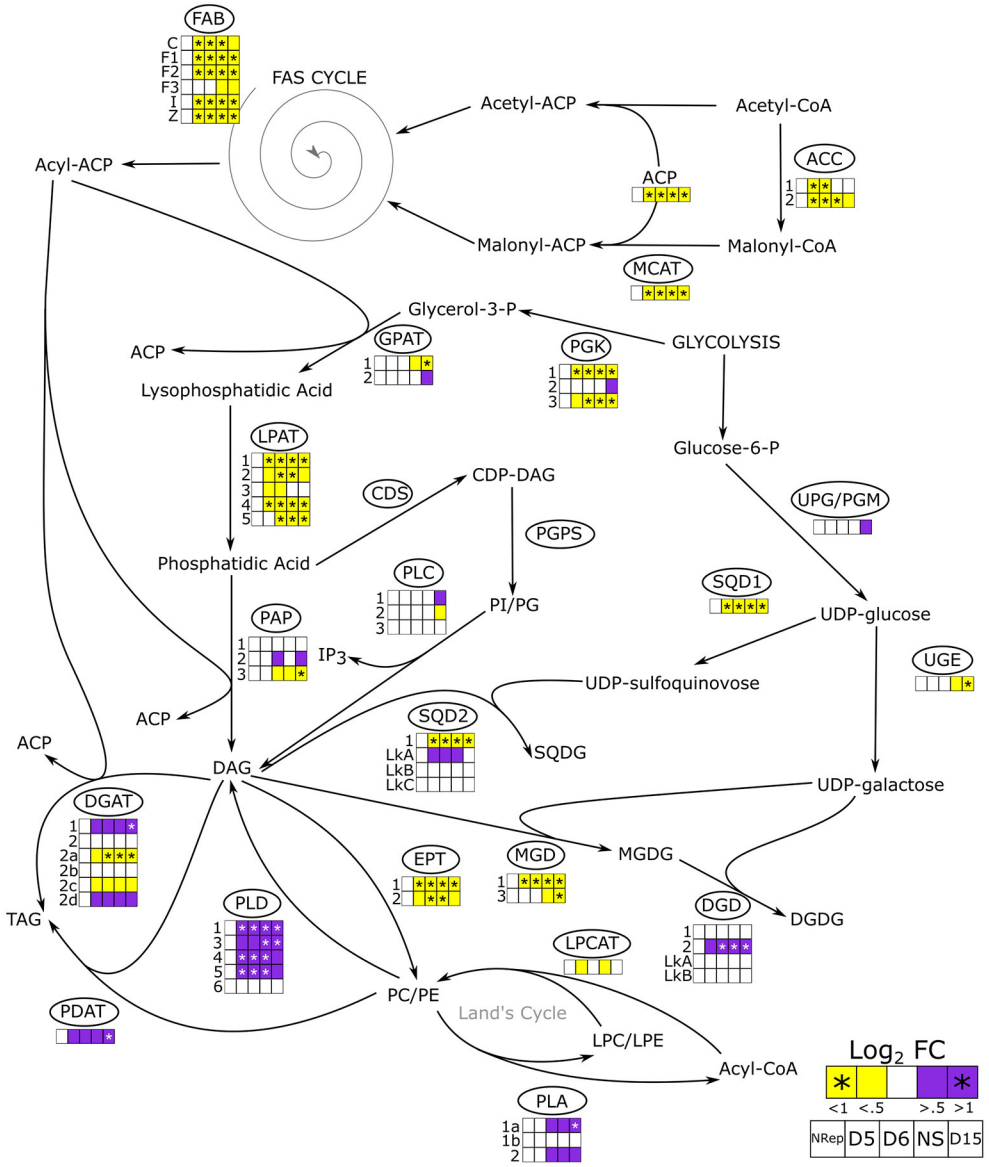


FIGURE 5 | *De novo* and acetyl-CoA independent lipid biosynthetic pathways in *P. tricornutum*. The overall change in gene expression at each phase. PDC, pyruvate dehydrogenase complex; ACC, acetyl-CoA carboxylase; MCAT, malonyl-CoA:ACP transacylase; FAB, fatty acid biosynthesis; PGK, phosphoglycerate kinase; GPAT, glycerol-3-phosphate: acyl-ACP acyltransferase; LPAT, lysophosphatidate acyltransferase; PAP, phosphatidate phosphatase; CDS, phosphatidate cytidylyltransferase; PGPS, CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase, PLC, phospholipase C; SQD1, UDP-sulfoquinovose synthase; SQD2, glycosyl transferase; UPG/PGM, UDP-glucose-pyrophosphorylase/phosphoglucomutase; UGE, UDP-galactose 4-epimerase; MGD, MGDG synthase; DGD, DGDG synthase; EPT, ethanolamine phosphotransferase; PLD, phospholipase D; PLA, phospholipase A; LPCAT, lysophosphatidylcholine acyltransferase; PDAT, phospholipid:diacylglycerol acyltransferase; DGAT, acyl CoA:diacylglycerol acyltransferase; ACP, acyl carrier protein. The heatmap is reported only when a change in gene expression is observed in at least one of the isoforms. Daily variations of gene expression are reported in **Supplementary Figure 11**.

tricornutum LDP1) (Wang et al., 2017; Lupette et al., 2019; Leyland et al., 2020). While PtLDP1 transcript abundance gradually decreased from the onset of the transition phase with a minimum at 15 DRNS, the StLDP encoding gene is overexpressed, mirroring the PtLDP1 expression pattern (**Supplementary Dataset 1**). Wang et al., (2017) reported that PtLDP1 upregulation resulted in an upregulation of DGAT2, GPAT, and FAS encoding genes, while in this experiment, an overall downregulation is observed for these genes (**Figure 5**). Moreover, a putative seipin homologous encoding gene (J47296) was mostly upregulated during the transition phase (**Supplementary Dataset 1**). The activity of seipins is important for LD biogenesis and maturation (Lu et al., 2017).

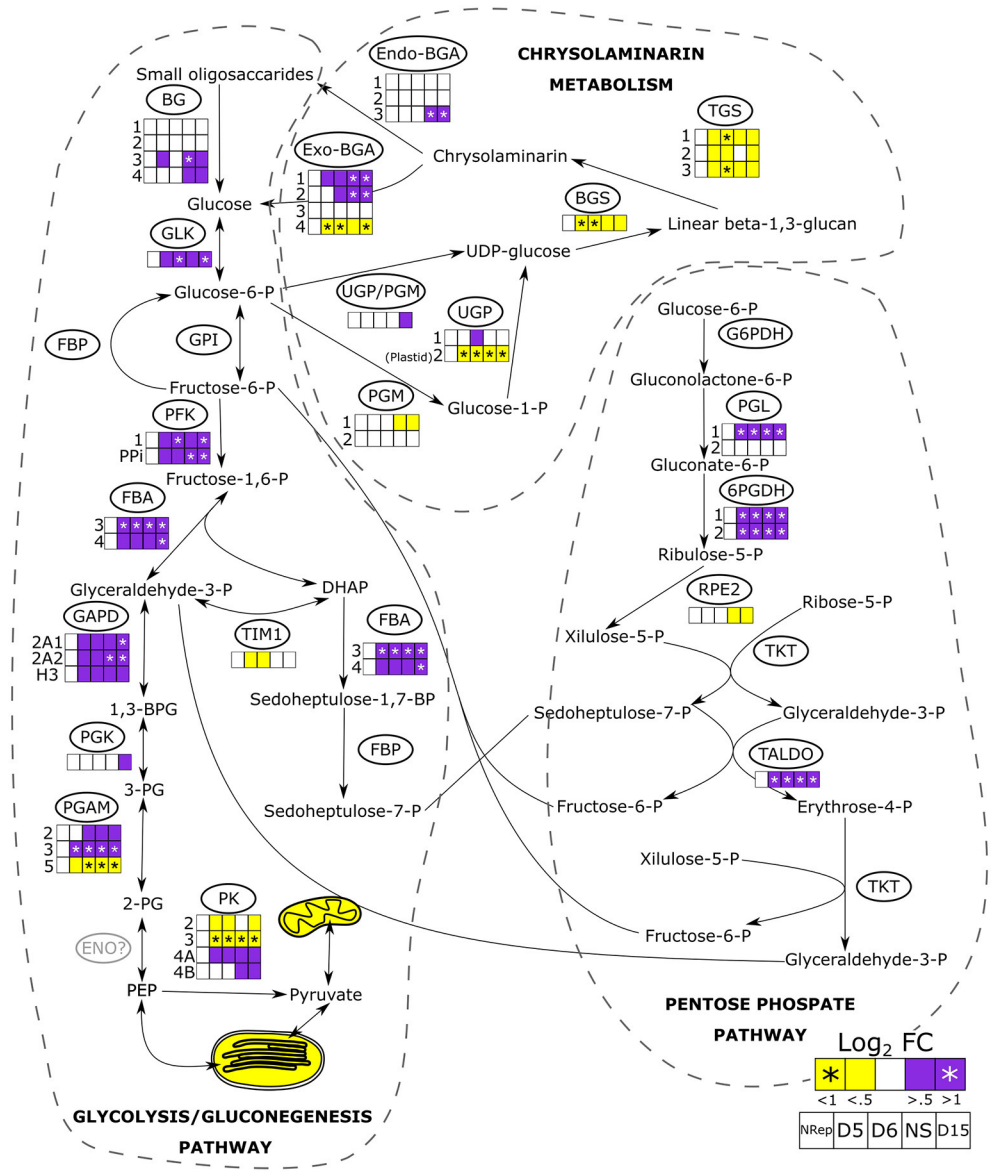


FIGURE 6 | Transcriptional changes of the genes involved in glycolysis/gluconeogenesis, pentose phosphate pathways, the synthesis and degradation of chrysolaminarin. Heatmaps are reported when changes in gene expression were observed in at least one of the isoforms. Mitochondrial and plastidial transcriptional changes are reported as overall change, the detailed pathway and transcript abundance change are reported in **Supplementary Figure 12**. Full information on the gene expression change during the time-course of the experiment is reported in **Supplementary Figure 13**. BG, β-glucosidase, GLK, glucokinase; GPI, glucose-6-phosphate isomerase; FBP, fructose-1,6-bisphosphatase; PFK, phosphofructokinase; FBA, fructose bisphosphate aldolase; GAPD, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGAM, phosphoglycerate mutase; ENO, enolase; PK, pyruvate kinase; TIM, triosephosphate isomerase; G6PDH, glucose 6-phosphate dehydrogenase; PGL, 6-phosphogluconolactonase, 6PGDH, phosphogluconate dehydrogenase; RPE2, ribulose-phosphate 3-epimerase; TKT, transketolase; TALDO, transaldolase; UPG/PGM, UDP-glucose-pyrophosphorylase/phosphoglucomutase; PGM, phosphoglucomutase; UGP, UTP–glucose-1-phosphate uridylyltransferase; BGS, 1,3-beta-glucan synthase; TGS, 1,6-β-transglycosylases; Exo-BGA, exo-β-glucanase; Endo-BGA, endo-β-glucanase.

The Central Metabolism

In diatoms, the cytosolic glycolysis pathway is partially duplicated in the mitochondria and the chloroplast (Kroth et al., 2008). When the N-ext became minimum at 5 DRNS (**Figure 1**), genes encoding for components of both organelle pathways started to be downregulated (**Figure 6**). While most of the glycolytic enzymes catalyzed also a reverse reaction in the gluconeogenesis, the phosphofructokinase (PFK) and pyruvate kinase (PK) are unidirectional enzymes, and the overexpression of their relative genes suggests an increase in C flux toward the catabolic pathway, even though two of the four PK genes were downregulated (**Figure 6**). Kroth et al. (2008) reported the presence of a plastidic enolase (ENO) and two other non-functional mitochondrial ones, but no cytosolic ENO has been described so far.

Other pathways parallel to glycolysis exist: one of these being the pentose phosphate pathway (PPP) (Kroth et al., 2008) and the other is an alternative to PPP which involves the xylulose-5-phosphate/fructose-6-phosphate phosphoketolase (XPK) (Fabris et al., 2012). The gene encoding this enzyme that is involved in these two pathways was upregulated only during the transition phase (**Supplementary Dataset 1**). Pyruvate, the product of glycolysis and alternative pathways, can then enter the tricarboxylic acid cycle (TCA) *via* two ways (**Supplementary Figure 13**): (1) through the conversion to acetyl-CoA by the activity of the pyruvate dehydrogenase complex (PDH), or (2) through the conversion to oxaloacetate by the action of the pyruvate carboxylase (PYC). The overall genes involved in the TCA were upregulated in NS, with an overall expression peak at 8 DRNS (**Supplementary Figure 13**).

β-glucans Metabolism

The β-1,3-glucans biosynthetic pathway (**Figure 6**) starts with the formation of glucose-1-phosphate catalyzed by the phosphoglucomutase (PGM), which transfers the phosphate of the glucose-6-phosphate from position 6 to 1. The conversion of glucose-1-phosphate to uridine diphosphate glucose (UDP-glucose) is catalyzed by the cytoplasmic UDP-glucose pyrophosphorylase (UGP) (Kroth et al., 2008). Then, UDP-glucose serves as a substrate for the linear polymerization catalyzed by the activity of β-1,3-glucan synthase (BGS, single gene). This reaction generates a linear β-1,3-glucan that can be further processed to β-1,3/1,6-glucan polymer, which is called chrysolaminarin. The last reaction is catalyzed by β-1,6-transglycosylases (TGS, 3 genes). The chrysolaminarin breakdown to glucose molecules is carried on by endo- and exo-β-glucanases (endo-BGS, exo-BGS) (**Figure 6**). Small oligosaccharides are further hydrolyzed to glucose by the activity of β-glucosidases (BG), glucose feeding the central metabolism.

An overall downregulation of the genes involved in chrysolaminarin biosynthesis was observed in the NS phase (**Figure 6**), while a significant increase in transcript level was observed for genes encoding the enzymes catalyzing the breakdown of β-1,3-glucans, which increase during the transition phase. They remained elevated all along the NS phase (**Figure 6**).

Photosynthesis and Carbon Fixation

Light Harvesting

Light harvesting complex (LHC) proteins are functional components of photosystems having different functions including light harvesting, photosystem regulation, and photoprotection (Bassi and Caffarri, 2001; Scarsini et al., 2019). In NS, most of the LHC genes were downregulated with few exceptions including *LHCX4* and *LHCR10*. These two transcripts, together with the high-light inducible protein 1 (*HLIP1*) transcript, were immediately overrepresented at the onset of the transition phase (**Figure 1**). Four other genes were upregulated in these conditions: *LHCF6*, *LHCF9,* and *LHCF15* were transiently upregulated during NS, whereas *HLIP1b* was transiently upregulated at the onset of NS. An overall strong downregulation was observed for the nuclear-encoded PSII reaction center (*psb*) genes involved in photosynthesis and photosynthetic electron transport chain (**Supplementary Dataset 1**). The only observed exception was the gene encoding the iron-sulfur clusters of the cytochrome b_6/f complex (PETC) that mediates the electron transfer between PSII and the PSI: it was upregulated during NS.

Photochemistry and Carbon Assimilation

The majority of the transcripts involved in photosynthesis were downregulated, in particular, a strong reduction was observed for the gene encoding the chloroplastic ferredoxin-NADP reductase (J23717) responsible for the photosynthetic production of NADPH, confirming the observations reported by Alipanah et al. (2015). In general, NS induced a strong reduction in transcript availability of most of the genes involved in the Calvin-Benson-Bassham cycle (**Supplementary Dataset 1**), while an overall increase in transcript abundance was observed for genes involved in the central PEP-pyruvate hub and for genes involved in the biochemical C4 cycle. High expression levels in NS were observed for the genes encoding the phospho*enol*pyruvate carboxylase 2 (PEPCase2), as well as its antagonist, the phospho*enol*pyruvate carboxykinase (PEPCK).

An increase in the transcript levels of genes encoding the malic enzyme (ME) and the 2-oxoglutarate: malate exchanger (EG02645) was observed in NS with two major expression peaks at 8 and 15 DRNS. While the malate exchanger encoded gene was upregulated, an overall downregulation was observed for the 4 plastidial triose phosphate transporters encoding genes (TPT1, TPT2, and TPT4a/b; Moog et al., 2015).

Most of the carbonic anhydrase (CA) genes were downregulated especially in NS with the sole exception of the CA1, which gene expression doubled at the onset of the NS phase (**Supplementary Dataset 1**).

Photorespiration

Huang et al. (2015) observed an increase in serine and glycine *de novo* biosynthesis *via* photorespiration, highlighting the potential importance of photorespiration in N redistribution in NS conditions. Our transcriptomic analysis showed an increase in several gene encoding for components of the photorespiration pathway. In fact, an increase in the transcript levels of genes encoding for serine/alanine:glyoxylate aminotransferase (J49601 and J43874) and the gene encoding for the glycine decarboxylase (EG02292) was observed (**Supplementary Dataset 1**). These two enzymes respectively converted the toxic glyoxylate in glycine than in serine, releasing one molecule of CO_2 and one molecule of ammonia (Kroth et al., 2008; Kunze and Hartig, 2013) (**Supplementary Figure 8**).

Photosynthetic and Photoprotective Pigment Synthesis

As expected, in NS, the *P. tricornutum* culture color faded from a pronounced brown-green to a very light yellow recalling the chlorotic phenotype observed in other photosynthetic organisms (Allen and Smith, 1969). The transcript abundance of most of the genes involved in the chlorophyll *a* biosynthetic pathway was decreased in this condition (**Supplementary Dataset 1**). The exception was observed for CHLC_2 and POR4, which were upregulated in NS. Simultaneously, genes encoding for the

carotenoid biosynthesis were downregulated in NS. Moreover, lowering the N availability triggered a reduction of the transcripts levels corresponding to two main key enzymes of the xanthophyll cycle, violaxanthin de/epoxidase, and zeaxanthin epoxidase (Scarsini et al., 2019).

Moreover, metabolite elucidation showed that 6 putative compounds belonging to the carotenoid biosynthetic pathway decreased in abundance during the time course of the experiment and, being in line with what was observed on a transcriptomic level (**Supplementary Dataset 1**).

DISCUSSION

The described analyses were performed in highly controlled growing conditions, allowing to observe the response of *P. tricornutum* to the sole NS while avoiding other limitations. The highly controlled turbidostat mode cultivation allows maintaining specific culture conditions for a long period free from the biases of the batch and fed-batch mods (Mcgeachy et al., 2019), while the addition of fresh medium is controlled by cell division rate. In fact, the latter is linearly proportional to nutrient availability and *vice versa* (Yaakob et al., 2021). In the absence of the desired modifications, such as the removal of nitrogen from the supplied medium, cells are continuously submitted to non-limiting nutrient availability. If N is removed from the supplied medium, it will be the only growth-limiting nutrient. Other nutrients, such as phosphorous, remained stable as they are uptaken by the cells with a proportional rate to the culture growth (Yaakob et al., 2021). In general, under turbidostat, each cell is submitted not only to constant nutrients, except for the desired removal, but also to light supply. By keeping the cell concentration sufficiently low, the self-shading effect will be minimized (Carvalho et al., 2011). This scenario cannot be reproduced in a batch culture where growing conditions continuously evolve. Not only does the nutrient availability per cell gradually decrease, but light distribution is affected by the increase in cell concentration as well. Continuous cultivation methods such as chemostat and turbidostat are important for experiment reproducibility, not only on a biochemical and physiological, but as well in a molecular level (Regenberg et al., 2006; Daran-Lapujade et al., 2009).

Intracellular Nitrogen Availability Guides the Passage Between Two Equilibria, NRep and NS

In nature, microalgae grow in very variable conditions and they may be submitted to reduced N availability that may eventually evolve into N starvation. As described above, the only modulated parameter is the N availability in the form of nitrate while all the other parameters are kept constant. The multidisciplinary approach allowed us to have a comprehensive view of the transition from NRep to NS. The broad picture of the metabolic reorientation, given by the PCA of transcriptomics, metabolomic and biochemical-physiological responses (**Figures 1**, **3**), was crucial to highlight that the DRNS induces a shift from an equilibrium to another. As expected, an equilibrium is observed in NRep conditions since cells are not subjected to limitation. Interestingly, a second equilibrium is observed when NS is reached. Cell metabolism has been reoriented and no qualitative changes were observed. Independently from the timepoint, cells in the NS phase possess the same metabolic instruction set. The observed changes during the NS phase are, thus, only quantitative (*e.g.,* increased neutral lipid content), while qualitative changes (*i.e.*, metabolic reorientation) have occurred during the transition phases.

The N-int appears to be the driving force that guides the shift between NRep and NS. In fact, cells remained in NRep phase until the minimum N-ext was reached and cells were forced to utilize the N-int. Division rate and N-int are closely interconnected (**Figure 1B**). In fact, the pool of N-int is split between the two daughter cells during cell division and the lower N-int will affect the cell division rate (Alipanah et al., 2015). An equilibrium is reached when the actual N availability is enough to maintain a minimum level of cell division rate. Even if the parameter guiding the NRep to NS transition is N-int, cells are not unaware of the changing environmental conditions before the trigger point. Indeed, 69 genes were differentially regulated at 4 DRNS including PLAAOx and CSAP encoding genes and one transcription factor of the heat shock family (HSF) (**Supplementary Dataset 1**).

Decreasing Nitrate Availability Triggers Nitrate Uptake and Storage Rather Than a Direct Utilization

The time shift of the decay of N-int compared with N-ext (**Figure 1**) is due to the presence of an intracellular N storage pool (Shoman, 2015). Aside from organic N, present in proteins and other N-containing molecules, nitrate can be directly stored in the cells. Mccarthy et al. (2017) reported that cells possess a remarkable ability to rapidly transport, assimilate, and also, presumably, store NO_3^- in vacuoles. The transient decrease of NR encoding transcript during the transition phase in contrast with increased transcript levels of genes encoding NO_3^- transporters of both plasmalemma and tonoplast, confirms the hypothesis of the potential storage of NO_3^- in the vacuoles (**Figure 4**). A temporary decrease in NR has been also observed by Longworth et al. (2016) within 72 h of NS.

Nitrate was the only N source in the medium, however, the effort of the cell to scavenge N from other N-containing molecules, such as free amino acids, is suggested by the upregulation of the gene encoding the PLAAOx. This enzyme oxidizes the aminoacidic groups to release ammonium which can be readily exploited by the cell as a source of N. The early overexpression of the PLAAOx and CSAP encoding genes (**Figure 4**), which has also been observed in *T. lutea* (Garnier et al., 2016), suggests a potential involvement of PLAAOx and CSAP in N sensing. Moreover, an alternative source of ammonium can be photorespiration which has already been observed as taking place in N-limited conditions (Huang et al., 2015). The increase in transcripts encoding, not only mitochondrial GOGAT but also serine/alanine:glyoxylate aminotransferase and glycine decarboxylase, leads to the

conclusion that in NS, a cellular effort in ammonia assimilation is observed following what was introduced by Remmers et al. (2018). Amidase activity is also contributing to ammonia scavenging, facing N deprivation (Alipanah et al., 2015).

Chloroplast Stand-by Mode as a Long-Term Cell Survival Mechanism

The passage to NS induces exposes cells to an energy imbalance between light availability and demand from cellular metabolism and growth (Remmers et al., 2018). The gradual development of the chlorotic phenotype is consistent with the decrease in transcript abundance of the majority of genes encoding LHCs, which is also reflected in the optimization of antenna size. On the other hand, the increased LHCX4 and LHCR10 transcript abundance suggests the need to increase the dissipation of excess absorbed energy. These two proteins belong to distinct families of LHCs, reported playing important roles in photoprotection (Lepetit et al., 2012; Collier Valle et al., 2014; Taddei et al., 2016; Buck et al., 2019). Both LHCX4 and LHCR10 encoding genes have been reported to be upregulated under nitrogen deprivation (Yang et al., 2013; Alipanah et al., 2015). While LHCR10 has been shown to be involved in photoprotection in response to the low light-to-high light transition of *P. tricornutum* (Collier Valle et al., 2014), LHCX4 does not appear to be involved in the fluorescence quenching photoprotection (Buck et al., 2019, 2021). Buck et al. (2021) suggested that one gene alone cannot contain all required regulatory *cis*-elements needed to respond to the multitude of environmental triggers and, thus, LHCX4 protein may play an integrative role between N starvation and photoprotection. Hence, the optimization of antenna size resulted in a decrease of the energy flow through the photosynthetic reaction centers, limiting photodamages and NADPH and ATP productions, the two key cofactors of the Calvin-Benson-Bassham cycle. In NS conditions, ribulose cannot be transformed into 3-phosphoglycerate but could enter the photorespiratory cycle (Sharkey et al., 1988; Bailleul et al., 2015). This is demonstrated by the increase of DIC (also observed by Valenzuela et al., 2012), the lack of increase in cell carbon quota (**Figure 1C**), the changes in expression of genes encoding Calvin-Benson-Bassham, and photorespiration components (**Supplementary Dataset 1**). This interpretation is further supported by the upregulation of the genes coding the enzymes involved in the photorespiratory pathway such as the alanine:glyoxylate aminotransferase (J49601, **Supplementary Dataset 1**).

A reduction of the plastidial activities follows a reduction of its energy content. Despite being exposed to low nitrogen availability for longer than 25 days, *P. tricornutum* was capable of a fast (within 24 hours) recovery of cell activities when a N source was reintroduced (**Supplementary Figure 6**). This fast recovery is incompatible with a complete shutdown and degradation of the chloroplast. These results confirmed previous observations made by Valenzuela et al. (2012). In short, NS turns chloroplast metabolism into a stand-by mode and, hence, the cell metabolism. In this scenario, the energy available in the chloroplast is not enough to keep it functioning and the mechanisms of organelle crosstalk need to be established as demonstrated by Bailleul et al. (2015). Indeed, the increase in the transcription levels for the gene encoding components of the malate valve (2-oxoglutarate:malate translocator and malic enzyme) suggests a connection between the different cellular pyruvate hubs and the energy transfer from the TCA to the plastid. Interestingly, the activation of the pyruvate hub (PEPCK, PEPC1, and PYC1) has been described as a default answer to the energy unbalance (Heydarizadeh et al., 2019), allowing the chloroplast stand-by mode as described in this work. This stand-by mode is reminiscent of the quiescent state described in *Chlamydomonas reinhardtii* (Tsai et al., 2018) and *Nannochloropsis oceanica* (Zienkiewicz et al., 2020). Regardless of their denomination, quiescent state, or standby mode, these mechanisms seemed highly conserved among microalgae and are crucial for cell survival in unfavorable conditions, such as those that may occur in nature. Deciphering these mechanisms may contribute to speeding up the development of microalgae as cell factories for the production of metabolites.

The Intricated Response of the Central Carbon Metabolism to Nitrogen Starvation

Central carbon metabolism plays an important role in carbon partitioning mainly toward the synthesis of the major organic macromolecule classes (*i.e.*, carbohydrates, proteins, and lipids), and plays a crucial role in response to environmental changes (Launay et al., 2020).

Different from TAG, the carbohydrate cellular quota is higher under NS but no continuous increase is observed (**Figure 1D**). In fact, Caballero et al. (2016) observed that *P. tricornutum* accumulates only 6% of cellular carbon in the form of chrysolaminarin in NS. In general, microalgae tend to break down storage carbohydrates in order to redirect carbon toward lipid biosynthesis (Hockin et al., 2012; Yang et al., 2013). Physiology and transcriptomic analyses showed a decrease in photosynthetic activity and carbon fixation through the Calvin-Benson-Bassham cycle, implying a decrease in *de novo* glucose generation and, thus, a reduced substrate for storage carbohydrates. The increase in carbohydrate cell quota, already reported in the literature (Caballero et al., 2016; Remmers et al., 2018), may suggest an increase in energy storage in the form of chrysolaminarin. However, according to the transcriptomic analysis, a decrease in chrysolaminarin biosynthesis and an increase in its breakdown could have occurred (**Figure 6**). The observed carbohydrate quota increase could be the effect of the active central metabolism suggested by the increase in transcript abundance of cytosolic glycolysis and TCA. The observed increase in carbohydrate quota is not continuous and occurred mostly during the transition phase. Thus, this increase is not due to an effort in carbohydrates storage but a carbon backbones remobilization and a central metabolism overactivity. The increase in transcript levels encoding two enzymes involved in gluconeogenesis, PEP carboxykinase (PEPCK) and pyruvate carboxylase 1 (PYC1), could suggest the neosynthesis of simple sugar and could explain the increase in carbohydrate cell quota. This hypothesis is corroborated by the observation of Matthijs et al. (2017) as an increase in maltotriose, maltose, and glucose in nitrogen-depleted *P. tricornutum* cells.

Remmers et al. (2018) stated that the central carbon metabolism response to NS is an intricate regulation of three oxidative pathways and gluconeogenesis. An increase in the transcript levels of the gene encoding components cytosolic classic glycolytic pathway is observed, suggesting an increase in glycolytic activity and confirming what was observed by Remmers et al. (2018) on both the transcriptomic and proteomic levels. Organelle counterparts are, on the other hand, downregulated. The increase in transcripts levels encoding components of the PPP, which produce NADPH (Kroth et al., 2008), has been observed. This pathway, together with chrysolaminarin degradation and glycolysis, suggested an increased effort of N starved cells to provide carbon fluxes through the TCA and balanced the reduced NADPH production from photosynthesis, as reported previously by Alipanah et al. (2015). The upregulation of genes encoding components of the TCA cycle also suggests an overactivity of answering the need of cells to compensate for the energy loss due to the reduced photosynthetic activity. Furthermore, TCA serves as a recycling of the carbon backbones derived from proteins and amino acids (Hockin et al., 2012). These catabolic activities produce oxaloacetate which can be converted in PEP by the activity of PEPCK or enter the malate shunt, contributing to the subcellular compartment energy balance (Selinski and Scheibe, 2019). Flügge et al. (2011) highlighted the crucial role of the PEP supply to plastids and its restriction, which may be eventually fatal for the plastid performance. Both the accumulation of TAG and carbohydrates, as well as protein degradation, are energy-requiring activities (Peth et al., 2013; Subramanian et al., 2013). Accelerating the central carbon metabolism allows the cells to both generate energy molecules and recycle carbon backbones during NS.

Lipids Accumulation During NR-to-NS Involves Intracellular Carbon Remobilization Rather Than CO_2 Fixation

The gradual accumulation of lipids and carbohydrates starts at the onset of the transition phase (**Figure 1D**). To explain this result, two alternative and non-exclusive hypotheses can be proposed: either the carbon required for the production of these molecules comes from the photosynthetic carbon fixation or/and from internal carbon reallocation. The data presented in this manuscript are in favor of the second hypothesis because all the genes involved in the FAS cycle are drastically downregulated, which would result in a strong reduction of the carbon flux through the *de novo* lipid biosynthesis pathway. The reduction in *de novo* FA biosynthesis has been previously observed in literature which supports the hypothesis of a lipid reallocation rather than biosynthesis (Valenzuela et al., 2012; Alipanah et al., 2015; Remmers et al., 2018). This reduction is attested by the strong increase in DIC, together with a decrease in photosynthesis (Φ_{II} and qP). Further proof of the reallocation of the carbon, rather than a fixation, is also explained by the very small percentage increase in cell carbon quota (**Figure 1C**), which does not fit with a *de novo* lipid biosynthesis deriving from carbon fixation. Moreover, the *de novo* fatty acid biosynthesis requires the acyl carrier protein (ACP), which expression is downregulated. Thus, in the absence of active carbon fixation, cells have no other choice but to convert pre-existing membrane lipids such as PC and PE. The choline and ethanolamine groups are removed by the action of the PLD, of which the majority of genes are upregulated in NS. The upregulation of genes encoding for the DGAT fits with an increase in TAG synthesis utilizing acyl groups, mostly generated by the breakdown of the phospholipids by the activity of the PLA, which encoding genes are upregulated. On the same frame, PDAT enzymes directly transfer an FA from the phospholipids to the DAG, while DGAT attaches a free FA group to DAG molecules. The same happens to the MGDG and DGDG, which are also a source of DAG and FA. Despite no gene encoding galactolipases have been so far identified in *P. tricornutum*, both transcriptomic and metabolomic data indicate an increase in galactolipid catabolism, thus, releasing fatty acids for TAG synthesis through lipid recycling.

CONCLUSIONS

The interdisciplinary approach used to study the kinetics of the response of *P. tricornutum* to a gradual reduction in N availability allowed us to observe that cells shifted from one metabolic status, in which energy is mainly exploited for cell division, to another one, in which cells division rate is reduced in favor of energy storage. The reduction in N availability forced the cells to exploit the intracellular N pool and, thus, induced a dismantling of non-essential N-containing cellular components. Division rate is reduced in response to low N and storage molecules, mainly neutral lipids, are accumulated. Plastids are the subcellular components mainly affected by NS and their activity is reduced to a "standby mode" with a lower carbon fixation level. The response resided in the intricated modulation of the central carbon metabolisms which is the flywheel of the cell, allowing to maintain a basal energy status under these sub-optimal conditions. In fact, even after a long N limitation, cells are capable of fast remobilization of the stored energy to recover their function. *P. tricornutum* cells responded to NS, and generally to environmental stresses, with a mechanism similar to quiescence, in which energy is stored and non-essential activities are drastically reduced.

DATA AVAILABILITY STATEMENT

The transcriptomic data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB47181 (https://www.ebi.ac.uk/ena/browser/view/PRJEB47181). The metabolic profiles are publicly available on the online infrastructure Ifremer Sextant catalogue of referential data from marine environments (https://doi.org/10.12770/38e00808-ffca-4266-943f-b691d3ca0bec). Data processing and statistical analysis of the metabolic profiles can be found at SANOE repository (https://doi.org/10.17882/84209). The data supporting the findings of this study are available from the corresponding author, Benoît Schoefs, upon request.

AUTHOR CONTRIBUTIONS

BS, JM, and MS: conceptualization. FM, JM, MS, and ST-R: data curation. BS, FM, JM, MS, and ST-R: formal analysis. BS, FM, HH, and JM: funding acquisition. BS, BV, JM, and MS: investigation. BS, BV, FM, JM, and MS: methodology. BS and JM: project administration. MS: software. BS, HH, and JM: supervision. BS, FM, JM, and MS: writing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Ewa Lukomska and Gaël Bougaran (Physiology and Biotechnology of Algae Laboratory, IFREMER, Nantes, France) for the C and N quantifications, as well as Manoela Sibat (Phycotoxins Laboratory, IFREMER, Nantes, France) for the technical help during metabolomic analyses. The authors are grateful to Le Mans University for the Ph.D. grant to MS and the post-doctorate position to ST-R. The Fondation BPO, the Institut Universitaire Mer et Littoral (IUML—FR 3473 CNRS), Le Mans University and the University Bretagne-Loire (UBL) are thanked for financial supports.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.760516/full#supplementary-material

REFERENCES

Abida, H., Dolch, L. J., Mei, C., Villanova, V., Conte, M., Block, M. A., et al. (2015). Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricornutum*. *Plant Physiol.* 167, 118–136. doi: 10.1104/pp.114.252395

Aderemi, A. O., Novais, S. C., Lemos, M. F. L., Alves, L. M., Hunter, C., and Pahl, O. (2018). Oxidative stress responses and cellular energy allocation changes in microalgae following exposure to widely used human antibiotics. *Aquatic Toxicol.* 203, 130–139. doi: 10.1016/j.aquatox.2018.08.008

Alipanah, L., Rohloff, J., Winge, P., Bones, A. M., and Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *J. Experim. Bot.* 66, 6281–6296. doi: 10.1093/jxb/erv340

Alipanah, L., Winge, P., Rohloff, J., Najafi, J., Brembu, T., and Bones, A. M. (2018). Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom *Phaeodactylum tricornutum*. *Plos ONE* 13:e0193335. doi: 10.1371/journal.pone.0193335

Allen, M. M., and Smith, A. J. (1969). Nitrogen chlorosis in blue-green algae. *Archiv. für Mikrobiologie* 69, 114–120. doi: 10.1007/BF00409755

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Apt, K. E., Kroth-Pancic, P. G., and Grossman, A. R. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. General Genet.* 252, 572–579. doi: 10.1007/BF02172403

Arras, K. O. (1998). *An Introduction to Error Propagation: Derivation, Meaning and Examples of Equation Cy= Fx Cx* FxT. PhD, Lausanne: École Polytechnique Fédérale de Lausanne.

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., Tanaka, A., et al. (2015). Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* 524, 366–U267. doi: 10.1038/nature14599

Bassi, R., and Caffarri, S. (2001). Lhc proteins and the regulation of photosynthetic light harvesting function by xanthophylls. *Photosynth. Res.* 64, 243–256. doi: 10.1023/A:1006409506272

Benoiston, A.-S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., et al. (2017). The evolution of diatoms and their biogeochemical functions. *Philosophic. Trans. Roy. Soc. B: Biol. Sci.* 372:20160397. doi: 10.1098/rstb.2016.0397

Blunt, J. W., Copp, B. R., Hu, W. P., Munro, M. H., Northcote, P. T., and Prinsep, M. R. (2008). Marine natural products. *Nat. Prod. Rep.* 25, 35–94. doi: 10.1039/b701534h

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochemistr.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Buck, J. M., Kroth, P. G., and Lepetit, B. (2021). Identification of sequence motifs in Lhcx proteins that confer qE-based photoprotection in the diatom *Phaeodactylum tricornutum*. *Plant J.* 21:15539. doi: 10.1111/tpj.15539

Buck, J. M., Sherman, J., Bartulos, C. R., Serif, M., Halder, M., Henkel, J., et al. (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. *Nat Commun.* 10:4167. doi: 10.1038/s41467-019-12043-6

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). *Phaeodactylum tricornutum*: A diatom cell factory. *Trends Biotechnol.* 38, 606–622. doi: 10.1016/j.tibtech.2019.12.023

Caballero, M. A., Jallet, D., Shi, L. B., Rithner, C., Zhang, Y., and Peers, G. (2016). Quantification of chrysolaminarin from the model diatom *Phaeodactylum tricornutum*. *Algal Res. Biomass Biofuels Bioproduct.* 20, 180–188. doi: 10.1016/j.algal.2016.10.008

Cagliari, A., Margis, R., Dos Santos Maraschin, F., Turchetto-Zolet, A. C., Loss, G., and Margis-Pinheiro, M. (2011). Biosynthesis of triacylglycerols (TAGs) in plants and algae. *Int. J. Plant Biol.* 2:e10. doi: 10.4081/pb.2011.e10

Carvalho, A., Meireles, L. A., and Malcata, F. (1998). Rapid spectrophotometric determination of nitrates and nitrites in marine aqueous culture media. *Analusis* 26, 347–351. doi: 10.1051/analusis:1998183

Carvalho, A. P., Silva, S. O., Baptista, J. M., and Malcata, F. X. (2011). Light requirements in microalgal photobioreactors: an overview of biophotonic aspects. *Appl. Microbiol. Biotechnol.* 89, 1275–1288. doi: 10.1007/s00253-010-3047-8

Collier Valle, K., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., et al. (2014). System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom *Phaeodactylum tricornutum*. *PlosONE* 9:e114211. doi: 10.1371/journal.pone.0114211

Courant, F. D. R., Martzolff, A., Rabin, G., Antignac, J.-P., Le Bizec, B., Giraudeau, P., et al. (2013). How metabolomics can contribute to bio-processes: A proof of concept study for biomarkers discovery in the context of nitrogen-starved microalgae grown in photobioreactors. *Metabolomics* 9, 1286–1300. doi: 10.1007/s11306-013-0532-y

Daran-Lapujade, P., Daran, J. M., Van Maris, A. J., De Winde, J. H., and Pronk, J. T. (2009). Chemostat-based micro-array analysis in baker's yeast. *Adv. Microb. Physiol.* 54, 257–311. doi: 10.1016/S0065-2911(08)00004-0

Dubois, M., Gilles, K. A., Hamilton, J. K., and Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Anal. Chemistr.* 28, 350–356. doi: 10.1021/ac60111a017

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo-Madrid, A. C., Abbriano, R. M., et al. (2020). Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids. *ACS Synthetic Biol.* 9, 598–612. doi: 10.1021/acssynbio.9b00455

Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A., and Baart, G. J. (2012). The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.* 70, 1004–1014. doi: 10.1111/j.1365-313X.2012.04941.x

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. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158

Flügge, U. I., Häusler, R. E., Ludewig, F., and Gierth, M. (2011). The role of transporters in supplying energy to plant plastids. *J. Experim. Bot.* 62, 2381–2392. doi: 10.1093/jxb/erq361

Flynn, K. J., and Butler, I. (1986). Nitrogen-sources for the growth of marine microalgae - role of dissolved free amino-acids. *Marine Ecol. Progr. Series* 34, 281–304. doi: 10.3354/meps034281

Garnier, M., Bougaran, G., Pavlovic, M., Berard, J.-B., Carrier, G., Charrier, A., et al. (2016). Use of a lipid rich strain reveals mechanisms of nitrogen limitation and carbon partitioning in the haptophyte *Tisochrysis lutea*. *Algal Res.* 20, 229–248. doi: 10.1016/j.algal.2016.10.017

Ge, F., Huang, W., Chen, Z., Zhang, C., Xiong, Q., Bowler, C., et al. (2014). Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum*. *Plant Cell* 26, 1681–1697. doi: 10.1105/tpc.114.124982

George, J., Kahlke, T., Abbriano, R. M., Kuzhiumparambil, U., Ralph, P. J., and Fabris, M. (2020). Metabolic engineering strategies in diatoms reveal unique phenotypes and genetic configurations with implications for algal genetics and synthetic biology. *Front. Bioeng. Biotechnol.* 8:513. doi: 10.3389/fbioe.2020.00513

Gnaiger, E. (1983). "Calculation of energetic and biochemical equivalents of respiratory oxygen consumption", in *Polarographic Oxygen Sensors*, eds E. Gnaiger and H. Forstner: Springer Berlin Heidelberg), 337–345.

Goold, H., Beisson, F., Peltier, G., and Li-Beisson, Y. (2015). Microalgal lipid droplets: composition, diversity, biogenesis and functions. *Plant Cell Rep.* 34, 545–555. doi: 10.1007/s00299-014-1711-7

Gordon, D., Merz, C. R., Gurke, S., and Schoefs, B. (2019). "Bubble farming: scalable microcosms for diatom biofuel and the next green revolution," in *Diatoms: Fundamentals and Applications,* eds J. Seckbach and R. Gordon. (Beverly, MA, USA.: Wiley-Scrivener), 583–659.

Gresham, D., and Hong, J. (2015). The functional basis of adaptive evolution in chemostats. *FEMS Microbiol. Rev.* 39, 2–16. doi: 10.1111/1574-6976.12082

Guillard, R. R., and Ryther, J. H. (1962). Studies of marine planktonic diatoms: I. Cyclotella nana and Detonula confervacea (Cleve). *Canad. J. Microbiol.* 8, 229–239. doi: 10.1139/m62-029

Heydarizadeh, P., Boureba, W., Zahedi, M., Huang, B., Moreau, B., Lukomska, E., et al. (2017). Response of CO_2-starved diatom *Phaeodactylum tricornutum* to light intensity transition. *Philosophic. Trans. Roy. Soc. B: Biol. Sci.* 372:20160396. doi: 10.1098/rstb.2016.0396

Heydarizadeh, P., Marchand, J., Chenais, B., Sabzalian, M. R., Zahedi, M., Moreau, B., et al. (2014). Functional investigations in diatoms need more than transcriptomic approach. *Diatom Res.* 29, 75–89. doi: 10.1080/0269249X.2014.883727

Heydarizadeh, P., Veidl, B., Huang, B., Lukomska, E., Wielgosz-Collin, G., Couzinet-Mossion, A., et al. (2019). Carbon orientation in the diatom *Phaeodactylum tricornutum*: the effects of carbon limitation and photon flux density. *Front. Plant Sci.* 10:471. doi: 10.3389/fpls.2019.00471

Hockin, N. L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012). The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol.* 158, 299–312. doi: 10.1104/pp.111.184333

Hu, H., and Pan, Y. (2020). "Electroporation transformation protocol for *Phaeodactylum tricornutum*," in *Electroporation Protocols: Microorganism, Mammalian System, and Nanodevice,* eds S. Li, L. Chang and J. Teissie. 3rd ed (New York, NY: Springer US), 163–167.

Huang, A. Y., Liu, L. X., Yang, C., and Wang, G. C. (2015). *Phaeodactylum tricornutum* photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. *Biotechnol. Biofuels* 8:73. doi: 10.1186/s13068-015-0256-5

Huang, B., Marchand, J., Thiriet-Rupert, S., Carrier, G., Saint-Jean, B., Lukomska, E., et al. (2019). Betaine lipid and neutral lipid production under nitrogen or phosphorus limitation in the marine microalga Tisochrysis lutea (Haptophyta). *Algal Res.* 40, 101–106. doi: 10.1016/j.algal.2019.101506

Karas, B. J., Diner, R. E., Lefebvre, S. C., Mcquaid, J., Phillips, A. P. R., and Noddings, C. M. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:7295. doi: 10.1038/ncomms7925

Kazamia, E., and Smith, A. G. (2014). Assessing the environmental sustainability of biofuels. *Trends Plant Sci.* 19, 615–618. doi: 10.1016/j.tplants.2014.08.001

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kroth, P. G., Chiovitti, A., Gruber, A., Martin-Jézéquel, V., Mock, T., Schnitzler Parker, M., et al. (2008). A model of carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *Plos ONE* 3:e1426. doi: 10.1371/journal.pone.0001426

Kunze, M., and Hartig, A. (2013). Permeability of the peroxisomal membrane: lessons from the glyoxylate cycle. *Front. Physiol.* 4:204. doi: 10.3389/fphys.2013.00204

Lands, W. E. (1960). Metabolism of glycerolipids. 2. The enzymatic acylation of lysolecithin. *J. Biol. Chemistr.* 235, 2233-2237. doi: 10.1016/S0021-9258(18)64604-6

Launay, H., Huang, W., Maberly, S. C., and Gontero, B. (2020). Regulation of carbon metabolism by environmental conditions: A Perspective from diatoms and other Chromalveolates. *Front. Plant Sci.* 11:1033. doi: 10.3389/fpls.2020.01033

Lepetit, B., Goss, R., Jakob, T., and Wilhelm, C. (2012). Molecular dynamics of the diatom thylakoid membrane under different light conditions. *Photosynth. Res.* 111, 245–257. doi: 10.1007/s11120-011-9633-5

Levy, B. S., and Patz, J. A. (2015). Climate change, human rights, and social justice. *Ann. Global Health* 81, 310–322. doi: 10.1016/j.aogh.2015.08.008

Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S., and Khozin-Goldberg, I. (2020). High resolution proteome of lipid droplets isolated from the pennate diatom *Phaeodactylum tricornutum* (Bacillariophyceae) strain pt4 provides mechanistic insights into complex intracellular coordination during nitrogen deprivation. *J. Phycol.* 56, 1642–1663. doi: 10.1111/jpy.13063

Liang, Y., Kong, F., Torres-Romero, I., Burlacot, A., Cuine, S., Légeret, B., et al. (2019). Branched-chain amino acid catabolism impacts triacylglycerol homeostasis in *Chlamydomonas reinhardtii*. *Plant Physiol.* 179, 1502–1514. doi: 10.1104/pp.18.01584

Lin, H. Y., and Lin, H. J. (2019). Polyamines in Microalgae: Something Borrowed, Something New. *Marine Drugs* 17:10001. doi: 10.3390/md17010001

Longworth, J., Wu, D. Y., Huete-Ortega, M., Wright, P. C., and Vaidyanathan, S. (2016). Proteome response of *Phaeodactylum tricornutum*, during lipid accumulation induced by nitrogen depletion. *Algal Res. Biomass Biofuels Bioprod.* 18, 213–224. doi: 10.1016/j.algal.2016.06.015

Love, M. I., Huber, W., and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome. Biol.* (2014) 15, 550. doi: 10.1186/s13059-014-0550-8

Lu, Y., Wang, X., Balamurugan, S., Yang, W. D., Liu, J. S., Dong, H. P., et al. (2017). Identification of a putative seipin ortholog involved in lipid accumulation in marine microalga *Phaeodactylum tricornutum*. *J. Appl. Phycol.* 29, 2821–2829. doi: 10.1007/s10811-017-1173-8

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., Schaller, H., et al. (2019). The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*. *Algal Res.* 38:101415. doi: 10.1016/j.algal.2019.101415

Matthijs, M., Fabris, M., Obata, T., Foubert, I., Franco-Zorrilla, J. M., Solano, R., et al. (2017). The transcription factor bZIP14 regulates the TCA cycle in the diatom *Phaeodactylum tricornutum*. *Embo J.* 36, 1559–1576. doi: 10.15252/embj.201696392

Mccarthy, J. K., Smith, S. R., Mccrow, J. P., Tan, M., Zheng, H., Beeri, K., et al. (2017). Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell* 29, 2047–2070. doi: 10.1105/tpc.16.00910

Mcgeachy, A. M., Meacham, Z. A., and Ingolia, N. T. (2019). An accessible continuous-culture turbidostat for pooled analysis of complex libraries. *ACS Synthetic Biol.* 8, 844–856. doi: 10.1021/acssynbio.8b00529

Mimouni, V., Ulmann, L., Pasquet, V., Mathieu, M., Picot, L., Bougaran, G., et al. (2012). The potential of microalgae for the production of bioactive molecules

of pharmaceutical interest. *Curr. Pharmaceutic. Biotechnol.* 13, 2733–2750. doi: 10.2174/138920112804724828

Moog, D., Rensing, S. A., Archibald, J. M., Maier, U. G., and Ullrich, K. K. (2015). Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. *Genom. Biol. Evol.* 7, 2955–2969. doi: 10.1093/gbe/evv190

Moss, R., Edmonds, J., Hibbard, K., Manning, M., Rose, S., Vuuren, D., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature* 463, 747–756. doi: 10.1038/nature08823

Pan, Y. F., Hu, F., Yu, C., Li, C. J., Huang, T., and Hu, H. H. (2020). Amino Acid Catabolism During Nitrogen Limitation in Phaeodactylum tricornutum. *Front. Plant Sci.* 11:589026. doi: 10.3389/fpls.2020.589026

Pan, Y. F., Yang, J., Gong, Y. M., Li, X. L., and Hu, H. H. (2017). 3-Hydroxyisobutyryl-CoA hydrolase involved in isoleucine catabolism regulates triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Philosophic. Trans. Roy. Soc. B-Biol. Sci.* 372:409. doi: 10.1098/rstb.2016.0409

Peth, A., Nathan, J. A., and Goldberg, A. L. (2013). The ATP costs and time required to degrade ubiquitinated proteins by the 26S proteasome. *J. Biol. Chemistr.* 288, 29215–29222. doi: 10.1074/jbc.M113.482570

Regenberg, B., Grotkjaer, T., Winther, O., Fausbøll, A., Akesson, M., Bro, C., et al. (2006). Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol.* 7:R107. doi: 10.1186/gb-2006-7-11-r107

Remmers, I. M., D'adamo, S., Martens, D. E., De Vos, R. C. H., Mumm, R., America, A. H. P., et al. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Res.* 35, 33–49. doi: 10.1016/j.algal.2018.08.012

Roháček, K., Soukupova, J., and Bartak, M. (2008). "Chlorophyll fluorescence: A wonderful tool to study plant physiology and plant stress," in *Plant Cell Organelles – Selected Topics,* ed. B. Schoefs. (Trivandrum: Research Signpost) 251–284.

Sayanova, O., Mimouni, V., Ulmann, L., Morant-Manceau, A., Pasquet, V., Schoefs, B., et al. (2017). Modulation of lipid biosynthesis by stress in diatoms. *Philosophic. Trans. Roy. Soc. B: Biol. Sci.* 372:1728. doi: 10.1098/rstb.2016.0407

Scarsini, M., Marchand, J., Manoylov, K., and Schoefs, B. (2019). "Photosynthesis in diatoms," in *Diatoms: Fundamentals and Applications,* eds. J. Seckbach and R. Gordon. (Beverly, MA, USA.: Wiley-Scrivener), 8.

Schoefs, B. (2004). Determination of pigments in vegetables. *J. Chromatograph. A* 1054, 217–226. doi: 10.1016/j.chroma.2004.05.105

Schoefs, B., Van De Vijver, B., Wetzel, C., and Ector, L. (2020). From diatom species identification to ecological and biotechnological applications. *Bot. Lett.* 167, 2–6. doi: 10.1080/23818107.2020.1719883

Selinski, J., and Scheibe, R. (2019). Malate valves: old shuttles with new perspectives. *Plant Biol.* 21, 21–30. doi: 10.1111/plb.12869

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9:3924. doi: 10.1038/s41467-018-06378-9

Sharkey, T. D., Berry, J. A., and Sage, R. F. (1988). Regulation of photosynthetic electron-transport in *Phaseolus vulgaris* L.; as determined by room-temperature chlorophyll *a* fluorescence. *Planta* 176, 415–424. doi: 10.1007/BF00395423

Sharma, A. K., Nymark, M., Sparstad, T., Bones, A. M., and Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation – Comparison with biolistic CRISPR/Cas9 transformation. *Scientific Rep.* 8:14401. doi: 10.1038/s41598-018-32342-0

Shoman, N. Y. (2015). The dynamics of the intracellular contents of carbon, nitrogen, and chlorophyll a under conditions of batch growth of the diatom *Phaeodactylum tricornutum* (Bohlin, 1897) at different light intensities. *Russian J. Marine Biol.* 41, 356–362. doi: 10.1134/S1063074015050132

Subramanian, S., Barry, A. N., Pieris, S., and Sayre, R. T. (2013). Comparative energetics and kinetics of autotrophic lipid and starch metabolism in chlorophytic microalgae: implications for biomass and biofuel production. *Biotechnol. Biofuels* 6:150. doi: 10.1186/1754-6834-6-150

Taddei, L., Stella, G. R., Rogato, A., Bailleul, B., Fortunato, A. E., Annunziata, R., et al. (2016). Multisignal control of expression of the LHCX protein family in the marine diatom *Phaeodactylum tricornutum*. *J. Experim. Bot.* 67, 3939–3951. doi: 10.1093/jxb/erw198

Thiriet-Rupert, S., Carrier, G., Trottier, C., Eveillard, D., Schoefs, B., Bougaran, G., et al. (2018). Identification of transcription factors involved in the phenotype of a domesticated oleaginous microalgae strain of *Tisochrysis lutea*. *Algal Res. Biomass Biofuels Bioproducts* 30, 59–72. doi: 10.1016/j.algal.2017.12.011

Tsai, C. H., Uygun, S., Roston, R., Shiu, S. H., and Benning, C. (2018). Recovery from N deprivation is a transcriptionally and functionally distinct state in chlamydomonas. *Plant Physiol.* 176, 2007–2023. doi: 10.1104/pp.17.01546

Valenzuela, J., Mazurie, A., Carlson, R. P., Gerlach, R., Cooksey, K. E., Peyton, B. M., et al. (2012). Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 5:40. doi: 10.1186/1754-6834-5-40

Vinayak, V., Manoylov, K. M., Gateau, H., Blanckaert, V., Herault, J., Pencreac'h, G., et al. (2015). Diatom milking: a review and new approaches. *Marine Drugs* 13, 2629–2665. doi: 10.3390/md13052629

Wang, X., Dong, H.-P., Wei, W., Balamurugan, S., Yang, W.-D., Liu, J.-S., et al. (2018). Dual expression of plastidial GPAT1 and LPAT1 regulates triacylglycerol production and the fatty acid profile in *Phaeodactylum tricornutum*. *Biotechnol. Biofuels* 11:318. doi: 10.1186/s13068-018-1317-3

Wang, X., Hao, T.-B., Balamurugan, S., Yang, W.-D., Liu, J.-S., Dong, H.-P., et al. (2017). A lipid droplet-associated protein involved in lipid droplet biogenesis and triacylglycerol accumulation in the oleaginous microalga *Phaeodactylum tricornutum*. *Algal Res.* 26, 215–224. doi: 10.1016/j.algal.2017.07.028

Wolfender, J.-L., Marti, G., Thomas, A., and Bertrand, S. (2015). Current approaches and challenges for the metabolite profiling of complex natural extracts. *J. Chromatograp. A* 1382, 136–164. doi: 10.1016/j.chroma.2014.10.091

Yaakob, M. A., Mohamed, R., Al-Gheethi, A., Aswathnarayana Gokare, R., and Ambati, R. R. (2021). Influence of nitrogen and phosphorus on microalgal growth, biomass, lipid, and fatty acid production: an overview. *Cells* 10:393. doi: 10.3390/cells10020393

Yang, Z.-K., Niu, Y. F., Ma, Y. H., Xue, J., Zhang, M. H., Yang, W. D., et al. (2013). Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnol. Biofuels* 6:67. doi: 10.1186/1754-6834-6-67

Zienkiewicz, A., Zienkiewicz, K., Poliner, E., Pulman, J. A., Du, Z.-Y., Stefano, G., et al. (2020). The microalga *Nannochloropsis* during transition from quiescence to autotrophy in response to nitrogen availability. *Plant Physiol.* 182, 819–839. doi: 10.1104/pp.19.00854

Zulu, N. N., Zienkiewicz, K., Vollheyde, K., and Feussner, I. (2018). Current trends to comprehend lipid metabolism in diatoms. *Progr. Lipid Res.* 70, 1–16. doi: 10.1016/j.plipres.2018.03.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Scarsini, Thiriet-Rupert, Veidl, Mondeguer, Hu, Marchand and Schoefs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fpls.2022.673113

Comparative Proteomic Analysis of the Diatom *Phaeodactylum tricornutum* Reveals New Insights Into Intra- and Extra-Cellular Protein Contents of Its Oval, Fusiform, and Triradiate Morphotypes

*Coralie Chuberre[1], Philippe Chan[2,3,4], Marie-Laure Walet-Balieu[1], François Thiébert[1], Carole Burel[1], Julie Hardouin[2,3,5], Bruno Gügi[1] and Muriel Bardor[1,6]**

[1] UNIROUEN, Laboratoire Glyco-MEV EA4358, Normandie Université, Rouen, France, [2] UNIROUEN, PISSARO Proteomic Facility, Institute for Research and Innovation in Biomedicine, Normandie Université, Mont-Saint-Aignan, France, [3] Normandie University, UNIROUEN, INSERM US 51, CNRS UAR 2026, HeRacLeS-PISSARO, Rouen, France, [4] UNIROUEN, Institute for Research and Innovation in Biomedicine, Normandie Université, Rouen, France, [5] Polymers, Biopolymers, Surface Laboratory, UMR 6270 CNRS, University of Rouen, Mont-Saint-Aignan, France, [6] Institut Universitaire de France, Paris, France

OPEN ACCESS

Edited by:
Justine Marchand, Le Mans Université, France

Reviewed by:
Richard Dorrell, École Normale Supérieure, France
Benoit Schoefs, Le Mans Université, France

***Correspondence:**
Muriel Bardor
muriel.bardor@univ-rouen.fr

Specialty section:
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

Received: *26 February 2021*
Accepted: *01 February 2022*
Published: *21 March 2022*

Citation:
Chuberre C, Chan P, Walet-Balieu M-L, Thiébert F, Burel C, Hardouin J, Gügi B and Bardor M (2022) Comparative Proteomic Analysis of the Diatom Phaeodactylum tricornutum Reveals New Insights Into Intra- and Extra-Cellular Protein Contents of Its Oval, Fusiform, and Triradiate Morphotypes. Front. Plant Sci. 13:673113. doi: 10.3389/fpls.2022.673113

Phaeodactylum tricornutum is an atypical diatom since it can display three main morphotypes: fusiform, triradiate, and oval. Such pleomorphism is possible thanks to an original metabolism, which is tightly regulated in order to acclimate to environmental conditions. Currently, studies dedicated to the comparison of each morphotype issued from one specific strain are scarce and little information is available regarding the physiological significance of this morphogenesis. In this study, we performed a comparative proteomic analysis of the three morphotypes from *P. tricornutum*. Cultures highly enriched in one dominant morphotype (fusiform, triradiate, or oval) of *P. tricornutum* Pt3 strain were used. Pairwise comparisons highlighted biological processes, which are up- and down-regulated in the oval (e.g., purine and cellular amino acid metabolism) and triradiate morphotypes (e.g., oxido-reduction and glycolytic processes) compared to the fusiform one used as a reference. Intersection analysis allowed us to identify the specific features of the oval morphotype. Results from this study confirmed previous transcriptomic RNA sequencing observation showing that the oval cells present a distinct metabolism with specific protein enrichment compared to fusiform and triradiate cells. Finally, the analysis of the secretome of each morphotype was also performed.

Keywords: diatom, *Phaeodactylum tricornutum*, microalgae, secretome, proteome, morphotype

INTRODUCTION

Phaeodactylum tricornutum is a pleomorphic raphid pennate diatom that exists naturally under three main morphotypes: fusiform, oval, and triradiate (Lewin et al., 1958; Borowitzka and Volcani, 1978; Johansen, 1991). The oval morphotype, which is preferentially benthic, possesses a raphe and organized silicified frustule unlike the fusiform and triradiate cells, which are more

planktonic species (Tesson et al., 2009; Vartanian et al., 2009). This pleomorphism, together with physiological and metabolic flexibilities, have been hypothesized to be responsible for the great adaptability of *P. tricornutum* to various environments (Rushforth et al., 1988; Gutenbrunner et al., 1994; De Martino et al., 2007; Butler et al., 2020; Falciatore et al., 2020; Song et al., 2020). Previous studies attempted to understand *P. tricornutum* polymorphism and demonstrated that such morphogenesis is not dependent on *P. tricornutum* genotype but can be induced by environmental factors (De Martino et al., 2007). Several works investigated recently the functional diversity of the ten ecotypes isolated from *P. tricornutum* that might also be a factor influencing the polymorphism (Rastogi et al., 2018, 2020; Song et al., 2020; Scarsini et al., 2021). The fusiform morphotype is the most frequent one in natural waters and *in vitro* cultures (Volcani, 1981; De Martino et al., 2007). In contrast, the triradiate are favored in an unstressed planktonic environment and preferentially developed with alkaline conditions (Johansen, 1991; De Martino et al., 2007; Bartual et al., 2008) whereas the oval cell growth rate increase in unfavorable and stress conditions (Borowitzka and Volcani, 1978; Johansen, 1991; De Martino et al., 2011). As morphotypes seem to be influenced by environmental conditions and as significant differences in the proteome regulation of *P. tricornutum* were reported in response to environmental conditions such as iron starvation (Allen et al., 2008), dark stress (Bai et al., 2016), or nitrogen deprivation (Yang et al., 2014; Longworth et al., 2016; Remmers et al., 2018), therefore, it is tempting to the hypothesis that proteomes might be different in the three morphotypes of *P. tricornutum*.

Previous comparative transcriptomic analyses of expressed sequence tags (EST) have suggested that the oval morphotype could be the most resistant form to stresses as this morphotype presents an ability to survive with limited nutrient availability and up-regulated genes encoding proteins involved in hyposalinity and cold stress responses (De Martino et al., 2007). Recently, *P. tricornutum* Pt3 strain was adapted to generate cultures enriched in one dominant morphotype: fusiform, triradiate, or oval. These cultures were used to run high-throughput RNA sequencing. The whole mRNA transcriptome of each morphotype was determined and pairwise comparisons highlighted biological processes and molecular functions, which are up- and down-regulated specifically (Ovide et al., 2018). In this previous study, less than 1% of genes were differentially expressed between the fusiform and the triradiate morphotypes whereas 22 and 29% were differentially expressed when comparing the oval morphotype *versus* the fusiform one and the oval morphotype *versus* the triradiate one, respectively (Ovide et al., 2018). Moreover, the metabolism of the oval cells was suggested to be specific compared to the other morphotypes (Ovide et al., 2018). Recently, Song et al. (2020) also reported that the oval cells synthesized a higher amount of proteins and pigments compared to the fusiform cells while the fusiform cell cultures accumulated lipids and carbohydrates. In addition, Galas and collaborators have shown that the oval Pt3 cells of *P. tricornutum* are secreted proteins more rapidly than the fusiform and triradiate Pt3 cells (Galas et al., 2021).

However, in this context, it remains unclear how the morphotype of *P. tricornutum* can impact the proteome. As previously suggested, it is essential to understand the interplay between the major players, e.g., genes (genomics), RNA (transcriptomics), proteins (proteomics), and metabolites in a cell type in order to decipher completely its cell biology (Heydarizadeh et al., 2014) and in the case of *P. tricornutum* its morphogenesis. Therefore, such a proteomic analysis will provide useful information to understand the pleiomorphism of the diatom *P. tricornutum* and could highlight morphotype-specific proteome signatures.

In this work, we performed a comparative proteomic analysis on the three morphotypes issued from the same *P. tricornutum* Pt3 strain. To reach this goal, cultures highly enriched in one dominant morphotype (fusiform, triradiate, or oval) of *P. tricornutum* Pt3 strain were used to prepare total protein extracts from cells representing the overall intracellular proteins. Then, label-free and iTRAQ® quantitative proteomics were applied to compare the proteome profile of cells issued from the three morphotypes. In addition, we also analyzed the proteins secreted in the culture medium of each morphotype by the label-free approach to characterize *P. tricornutum* secretomes of each morphotype.

MATERIALS AND METHODS

Experimental Design and Setup

Diatom cells of the Pt3 strain (CCAP 1052/1B; CCMP 2558), initially derived as a subclonal culture of Pt2 in Plymouth (De Martino et al., 2007), were grown at 19°C under a 16 h/8 h light/night cycle. Fusiform and triradiate cells were cultivated in sterilized 100% natural seawater (SW) (33.3 g/L, Instant Ocean, Aquarium System, Sarrebourg, France) and oval cells were cultivated in 10% SW (3.3 g/L). SW was then complemented with a nutrient medium (Conway 1 mL/L) and a metasilicate sodium solution (80 mg/L) (Ovide et al., 2018). Typical confocal and transmission electron microscopy (TEM) images of each morphotype can be found in Ovide et al. (2018) and Galas et al. (2021). In this work, the cultures were non-axenic. Particular attention has been paid to minimizing the presence of cellular debris in the culture medium. To do so, cells were pre-cultured (250 mL) twice in flasks under orbital agitation (150 rpm) on an IKA KS 260 Basic shaker (Sigma, St Quentin Fallavier, France) for 4 days in order to reach an exponential growth phase. Then, the cells were spun down and washed with fresh medium (SW) in order to remove cellular debris. Finally, the washed cell pellets were used to run a 1 L bioreactor using the culture conditions as previously described. Five biological replicates were performed and enriched cultures in one specific and dominant morphotype, as described in **Table 1**, reflecting the homogeneity of the samples.

Cell type proportions were estimated using an optical microscope associated with a manual cell count using a Nageotte cell ($n = 5$).

In order to establish the different proteomes from the three main morphotypes of *P. tricornutum*, intracellular proteins and proteins secreted within the culture media, later called secretome in this paper, were independently extracted and analyzed.

TABLE 1 | Relative percentage of each specific morphotype in *P. tricornutum* enriched culture.

Dominant morphotype	Morphotype enrichment* of *P. tricornutum* cultures
Fusiform	90 ± 1.2%
Triradiate	77 ± 0.5%
Oval	98 ± 0.4%

**Enrichments are expressed in specific morphotype cells number per 100 of total cells. Means were calculated over five biological replicates ± SE.*

Intracellular Proteome Extraction

Total proteins were extracted from Pt3 cultures enriched in one specific and dominant morphotype. Microalgae cells from the different cultures were recovered by centrifugation at 4,500 × *g* for 5 min.

Cell pellets were washed twice with 10% SW in order to decrease salt concentration. For each culture, a pellet of approximately 1.10^8 cells was harvested and immediately re-suspended in 500 μL of D2R2 protein extraction buffer {7 M urea, 2 M thiourea, 2 mM tributyl phosphate (TBP), 0.5% 3-(4-heptyl) phenyl-3-hydroxypropyldimethylammoniopropanesulfonate (C7BzO), and 2% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (Chaps)}. The cell suspension was stored at −80°C until further use. For protein extraction, the cell suspension was transferred into a 2 mL tube of lysing matrix E (MP Biomedicals, Fisher Scientific, Illkirch, France), and cells were lysed by 4 runs of 30 s at 6.5 $m.s^{-1}$ with the FastPrep®-24 high-speed benchtop homogenizer (MP Biomedicals®). Cell lysates were centrifuged at 10,000 × *g* for 10 min in order to remove cellular debris. The supernatants containing proteins were collected. The remaining pellet was re-suspended in 500 μL of D2R2 protein buffer and extracted once more. Finally, the supernatants were pooled in a unique fraction called IP for intracellular proteins.

Secreted Proteins From the Culture Medium

For the analysis of the secretomes (proteins secreted in the culture medium), a volume of 28 mL of each culture was centrifuged at 100,000 × *g* for 1 h at 4°C allowing removal of residual cell debris and eventual extracellular vesicles that could be present in the culture media. The supernatant was concentrated in the Pierce™ protein Concentrator 3K MWCO (Fisher Scientific, Illkirch, France) by centrifugation at 4,000 × *g* for 30 min and washed twice with deionized water. Therefore, all peptides below 3 kDa were lost during this step and only higher proteins have been analyzed. Collected supernatants containing secreted proteins, later called SP, were lyophilized, and then resuspended in deionized water prior to proteomic analysis.

Polyacrylamide Gel Electrophoresis and Protein Staining

A sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed to monitor the quality of the protein extraction. Samples were loaded on a precast Bolt™ 4–12% Bis-Tris Plus Gel (Invitrogen, Fisher Scientific, Illkirch, France). after denaturation in 5X Laemmli buffer for 5 min at 100°C. Electrophoresis was performed at room temperature for approximately 45 min using a constant voltage of 200 V in a 1X solution of NuPAGE MOPS SDS running buffer (Invitrogen, Fisher Scientific, Illkirch, France) until the dye front reached the end of the gel. Proteins were stained using an home-made Coomassie blue R-250 solution (30% ethanol, 10% acetic acid, and 0.02% Coomassie R-250).

Proteomic Sampling and Preparation

Collected supernatants called IP (640 μL) and SP (100 μl) were mixed with 5X Laemmli Buffer (0.312 M Tris (sans HCl) pH 6.8, 50% v/v glycerol, 10% w/v SDS, 5% v/v B-mercaptoethanol, and 0.25% bromophenol blue) and heated at 100°C for 10 min. Samples (600 μL for IP and 80 μL for SP) were loaded on a 7% bis-acrylamide gel. Electrophoresis was performed at 10°C for approximately 3 h using a constant amperage of 20 mA in a Tris-Glycine buffer (1.4% w/v glycine, 0.3% w/v Tris base, and 0.1% w/v SDS). Gels were washed twice in deionized water and incubated for 30 s in the Coomassie blue R-250 solution under shaking. The protein band just above the migration front was cut off and stored in a mix of 30% ethanol and 10% acetic acid until further analysis. Excised protein bands either from IP or SP have been reduced in 5 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP, Sigma, Saint-Quentin Fallavier, France) in 100 mM TEAB at 56°C for 1 h as previously described in Young et al. (2015). Then, reduced cysteine residues were blocked using 55 mM iodoacetamide (Sigma, Saint Quentin Fallavier, France) at room temperature for 20 min before performing in-gel trypsin digestion according to the manufacturer's instructions (Promega, Charbonnières-Les-Bains, France). The peptides were extracted once with formic acid 1%, twice with 100% acetonitrile/5% formic acid (v/v) and combined fractions dried using a vacuum concentrator (Young et al., 2015). Finally, peptide samples were resuspended in 100 μl of deionized water. Moreover, 10 μl of IP samples were used for label-free proteomic analysis, and 30 μl of IP were subjected to iTRAQ® labeling according to the manufacturer's recommendations (AB Sciex SAS, Villebon Sur Yvette, France). The three different morphotype samples: fusiform, oval, and triradiate were respectively labeled, overnight at room temperature, by iTRAQ® 117/115/116 in a 3Plex experimental design. Samples from each biological replicate were finally pooled and dried-down before lnanoliquid chromatography coupled with electrospray ionization tandem mass spectrometry (nanoLC-MS/MS) analysis.

NanoLC-ESI-MS/MS Analysis

Before running the nanoLC-ESI-MS/MS analysis, all peptide samples were resuspended in 5% (v/v) acetonitrile and 0.1% (v/v) formic acid: IP label-free (15 μL), IP iTRAQ® labeling (100 μL) and SP (12 μL). Two microliters of each sample was then analyzed on the Q-Exactive Plus (Thermo Scientific, Les Ulis, France) equipped with a nanoESI source. Peptides were loaded onto an enrichment column [C18 Pepmap100 (5 mm × 300 μm i.d., 5 μm, 100 Å), Thermo Scientific, Les Ulis, France] and

separated on an EASY-spray column [(50 cm × 0.075 mm i.d., 3 μm, 100 Å), Thermo Scientific, Les Ulis, France] with a flow rate of 300 nL.min^{-1}. The mobile phase was composed of H_2O/0.1% formic acid (buffer A) and acetonitrile/H_2O/0.1% formic acid (80/20) (buffer B). The elution gradient duration was 120 min following different steps: 0–84 min, 2–35% B; 84–94 min, 35–90% B; 94–105 min, 90% B; 106–120 min, 2% B. The temperature of the column was set at 40°C. The mass spectrometer acquisition parameters were: 100 ms maximum injection time, 1.6 kV capillary voltage, 275°C capillary temperature, full scan MS m/z @ 400–1,800 with a resolution of 70,000 in MS and 17,500 in MS/MS. The 10 most intense ions (Top 10) were selected and then fragmented with nitrogen as a collision gas (normalized collision energy set to 27 and 38 eV for iTRAQ®). All spectra obtained were exported in "raw" format that was used for data analysis.

Proteomic Data Analysis

Label-free and iTRAQ® analyses were performed to identify and quantify the differential expression levels of intracellular proteins. Label-free approach analysis was performed to identify the secretomes. For intracellular proteins, datasets were normalized, counted prior to identification using Mascot[1] and Progenesis software as compared to the *P. tricornutum* Uniprot database[2].

For the quantification and the identification of the differentially regulated intracellular proteins of each morphotype, Progenesis liquid chromatography-mass spectrometry (LC-MS) software (Nonlinear Dynamics, version 4.1[3]) was used for label-free peptide approach as previously described (Fréret et al., 2013). Automatic alignment was set to perform two by two comparisons between samples in order to align the LC-MS runs to account for retention time drifts. A minimum of 80% alignment score was required for further analysis. After alignment, statistical analysis was performed with one-way ANOVA calculations. To highlight differentially expressed peptides between groups (triradiate or oval morphotype against fusiform used as a reference), an ANOVA p-value ≤ 0.05 was required. The peak list containing the differential expressed peptides were then used for identification using Mascot (Matrix Science, version 2.5, Boston, MA, United States) with the following parameters: enzyme specificity, trypsin; one missed cleavage permitted; variable modifications, carbamidomethylation (C); oxidation (M), pyro glu from E and Q; monoisotopic precursor mass tolerance: 5 ppm; product mass tolerance: 0.02 Da against the *P. tricornutum* Uniprot reference proteome which was cleaned by removing pseudogenes. A positive match was considered when it was ranked among the first positions and presented a score with a significant threshold of $p < 0.05$ and with a false discovery rate (FRD) below 1. Mascot search results were imported back into Progenesis for differential expression and only proteins with an ANOVA p-value ≤ 0.05 and identified by ≥ 2 peptides were retained. The quality of the data was checked by principal component analysis (PCA) (**Supplementary Figure 1**).

For protein identification in SP and iTRAQ® labeling IP samples, peak lists were extracted (merge MS^n scans with the same precursor at ±30 s retention time window and ±50 ppm mass tolerance) and compared with specific databases by using the PEAKS studio 7.5 proteomics workbench (Bioinformatics Solutions Inc., Waterloo, Canada, build 20150615). The searches were performed with the following specific parameters: enzyme specificity, trypsin; three missed cleavages permitted; fixed modification, carbamidomethylation (C); variable modifications, oxidation (M), pyro-glu from E and Q; monoisotopic; mass tolerance for precursor ions, 5 ppm; mass tolerance for fragment ions, 0.02 Da; MS scan mode FT-ICR/Orbitrap; MS/MS scan mode, Linear ion Trap; Fragmentation mode, high energy CID; databases, *P. tricornutum* databases (UniProt Phaeodactylum + AND + tricornutum + AND + %22%28strain + CCAP + 1055%2F1%29%22 and Ensembl release 41 ASM15095v2.pep.03092018 [see test footnote 2]). Database search results were used for quantitative analysis with PEAKS Q and iTRAQ-4plex as selected methods. Only significant hits with a false discovery rate (FDR ≤ 1) for peptide and protein cut off (–logP ≥ 20 and unique peptides ≥ 2) and fold change > 1.2 (iTRAQ® hit map) were considered. On this dataset, less than 0.5% of peptides presented a miss-cleavage greater than 1.

Functional Annotation

The steps implemented to perform functional annotation are summarized in **Figure 1**. They lead to a functional Gene Ontology (GO) analysis: GO annotation and GO enrichment were performed using Blast2GO software version 3.3.536[4]). Protein sequences in FASTA were obtained with dataset *P. tricornutum* ASM15095v2 (Ensembl protist genes 47[5]). Homology searches have been launched against NCBI nr Database (QBlast). Then, default parameters were used to map, annotate, and run InterPro scan[6]. Pie charts were done using the Multi-Level option in Blast2GO considering the biological process (Ovide et al., 2018). Intersection analysis was performed using the Venny version 2.1 (Oliveros, 2007).

Pairwise comparison of the morphotype's intracellular proteomes was handled by considering that the fusiform morphotype was the reference as it is described to represent the most common morphotype (De Martino et al., 2007, 2011; Ovide et al., 2018).

Overrepresentation Analysis

An overrepresentation test was performed using the PANTHER classification system (PANTHER version 15[7]; Mi et al., 2013). UniProtKB: UniProt accession sublists of selected proteins either from the fusiform, oval or triradiate morphotypes were loaded on PANTHER and compared to IDs from Reference Proteome Genome (*P. tricornutum* PHATC) using GO-slim PANTHER biological process. Fisher test and FDR correction were applied. The outcomes summarized in a Microsoft® Excel® spreadsheet,

[1]https://www.matrixscience.com/, Boston, United States.

[2]https://www.uniprot.org/proteomes/UP000000759

[3]Waters corporation, https://www.nonlinear.com/

[4]https://www.Blast2GO.com

[5]http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index

[6]https://www.ebi.ac.uk/interpro/search/sequence/

[7]http://www.pantherdb.org

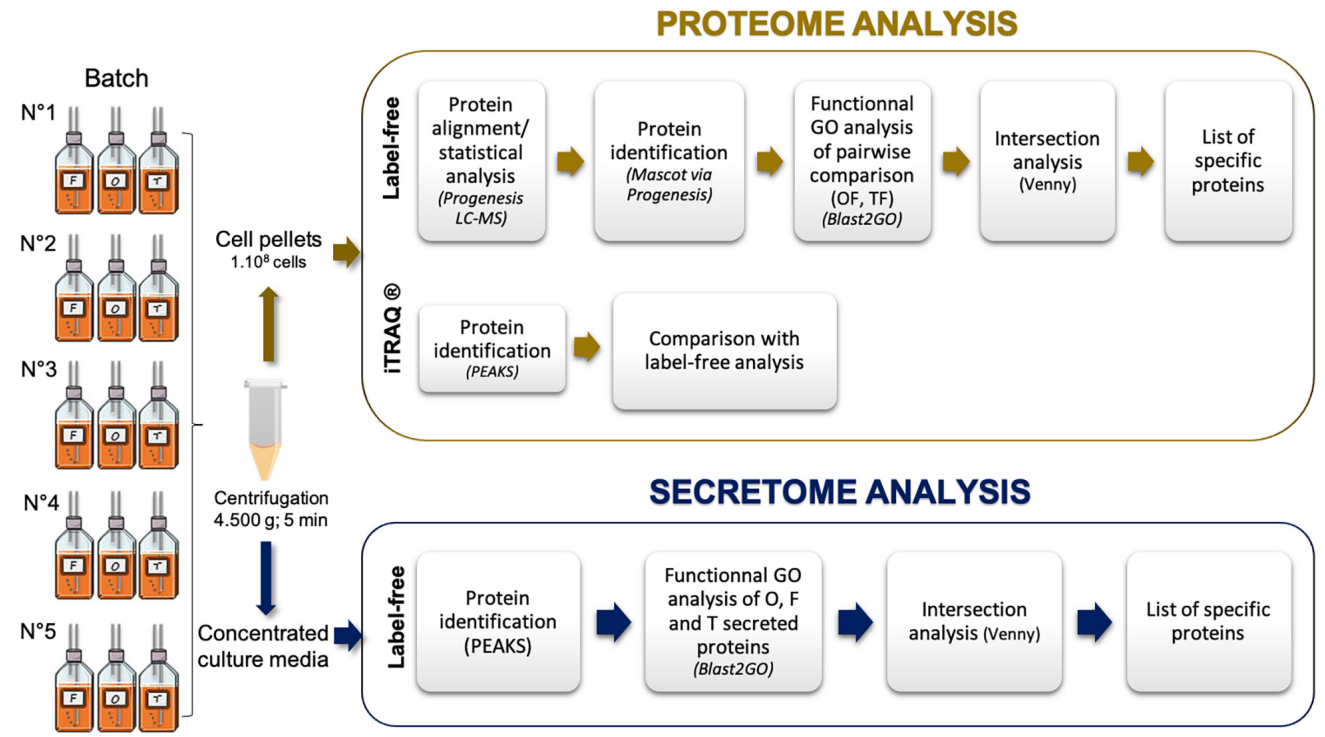


FIGURE 1 | Experimental design used in the current study to identify the intracellular proteomes and secreted proteins of the three main morphotypes of *Phaeodactylum tricornutum*. F, fusiform morphotype; O, oval morphotype; T, triradiate morphotype. Five biological replicates have been performed and summed for the data analyses.

Microsoft, United States were used to draw the related final figures.

Prediction of Signal Peptide and *N*-Glycosylation Sites on Secreted Proteins

The prediction of a potential signal peptide on the secreted proteins has been analyzed by the SignalP version 5[8] (Almagro Armenteros et al., 2019). The probability was considered correct when the score was greater than 0.5. The online tool NetNGlyc[9] has been used to predict potential *N*-glycosylation sites of the secreted proteins. This dataset was then compared to the ones reported in the studies of Rastogi et al. (2018) and Dorrell et al. (2021).

RESULTS

Phaeodactylum tricornutum cell cultures enriched in one dominant morphotype: Fusiform, Oval, and Triradiate were used in this work. Such cultures have been characterized recently using TEM and confocal microscopies (Ovide et al., 2018; Galas et al., 2021). Especially, the latest study demonstrated that the three morphotypes share similarities in terms of organelle localization (nucleus, mitochondria, F-actin cortex, intracellular network, etc.). In contrast, compared to fusiform and triradiate cells, oval cells spontaneously release proteins more rapidly, underlying a more rapid protein secretion in the oval morphotype (Galas et al., 2021). In the present study, we analyzed through proteomic the intracellular proteins and secreted proteins from the different morphotypes issues from the same Pt3 strain (**Figure 1**).

Differences in Intracellular Protein Productions Exist Between the Three Pt3 *Phaeodactylum tricornutum* Morphotypes

The intracellular protein profile from the oval morphotype (O) was compared to the one of the fusiform (F) morphotypes used as a reference (O vs. F, noted OF). When comparing OF, a total of 691 proteins were identified. As represented in the volcano plot shown in **Figure 2A**, 61 proteins were statistically differentially produced at a log fold change greater than −2 or 2 in the O morphotype compared to the F morphotype. Among them, 38 proteins were down-regulated whereas 23 were up-regulated respectively. In contrast, when comparing the Triradiate (T) morphotype to the F one (T vs. F noted TF), only 50 proteins were differentially produced with a log fold change comprised between −2 and 2 (**Figure 2B**). No proteins were up or down regulated at a log fold change higher than −2 or 2 in the

[8]https://services.healthtech.dtu.dk/service.php?SignalP-5.0
[9]https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0

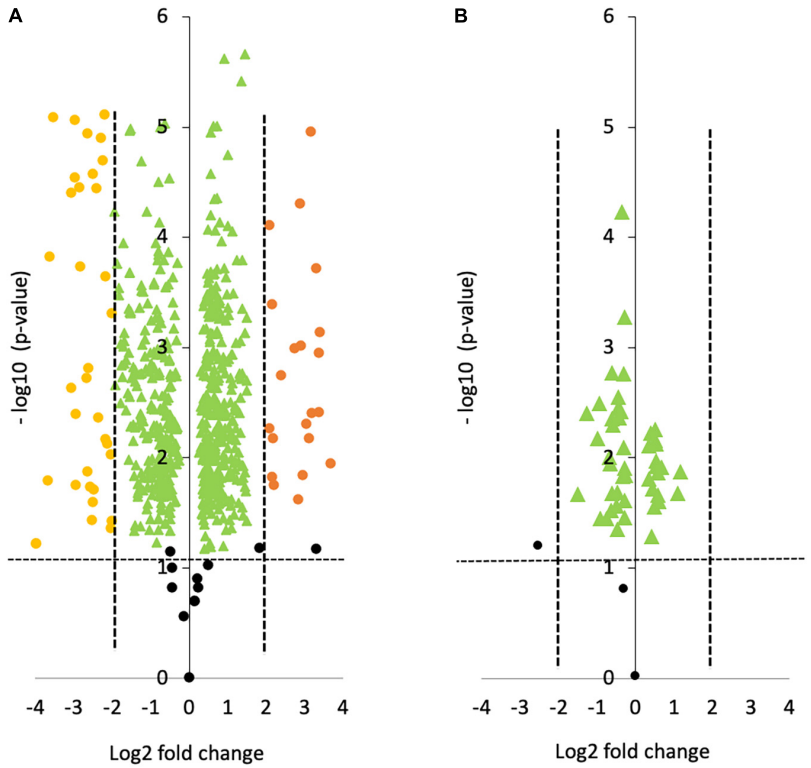


FIGURE 2 | Volcano plots representing the pairwise comparison of **(A)** Oval vs. Fusiform (OF); **(B)** Triradiate vs. Fusiform (TF). The volcano plots show a change in protein expression (*t*-test *p*-value) between **(A)** oval and fusiform cells, **(B)** triradiate and fusiform cells. Fusiform cells are considered as the reference. Orange and yellow dots represent up-regulated and down-regulated proteins at a log fold change upper or lower than 2, respectively (p-value ≤ 0.05). Green dots represent protein that has a fold change comprised between –2 and 2. Black dots represent the non-differentially expressed proteins.

TF comparison. These results suggest that there is not much difference in the intracellular protein profile of TF.

Identification of these differential proteins was performed using Blast2GO. This software allows the characterization and classification of proteins into biological processes, molecular and cellular functions according to GO annotation. At this stage, non-significantly differentially produced proteins (p-value > 0.05) identified from the proteomic label-free analysis were removed (11 proteins). Thus, the Blast2GO analysis was run on 680 proteins resulting from the OF pairwise comparison and 48 proteins for the TF pairwise comparison (p-value ≤ 0.05). From this analysis, about 84% of the intracellular proteins from the OF and 94% of intracellular proteins from the TF pairwise comparison have been annotated with GO terms thus corresponding to a total of 569 and 45 proteins respectively. Repartition of the identified proteins was performed based on their involvement in specific biological processes. A Pie chart representing the OF biological processes (**Figure 3A**) showed that oxidation-reduction (16%), cellular amino acid (13%), carbohydrate (12%), translation (10%), and energy (11%) processes represent overall 62% of the differentially produced proteins. Transport and phosphorylation processes represented 7 and 8%, respectively. In contrast, purine nucleotide metabolic pathway and ribonucleotide metabolic pathways represent 2 and 1% respectively. Other biological processes such as small biosynthetic molecules, heterocycle, cyclic compounds, and monocarboxylic acid metabolic pathways represent less than 5% for each category. A GO category dedicated to biological regulation also emerged at 2%.

To further understand the specific regulation between *P. tricornutum* morphotype, up- and down-regulated proteins were subjected to GO analysis separately. In brief, 55% of the 569 annotated proteins with a GO term are up-regulated in the O morphotype and 45% are down-regulated in comparison to the F morphotype. Among the up-regulated proteins (**Figure 3B**), 13% are related to the cellular amino acid process; 9% correspond to the carbohydrate derivative metabolic process, 9% to phosphorylation, and 8% to the carbohydrate metabolic process. Purine nucleotide metabolic and nucleic acid metabolic processes represent 2% respectively whereas cellular protein metabolic and gene expression processes represent 4% (**Figure 3B**). In contrast, proteins associated with the synthesis of precursor metabolites and to energy processes (21%), to translation (18%), proton transmembrane transport (12%), carboxylic acid metabolism (8%) and phosphate-containing compound metabolism (8%) processes are down-regulated in O cells (**Figure 3C**). Particular attention was paid to the oxidation-reduction GO term as proteins involved in this biological process are either up-regulated (14%) or down-regulated (23%) in the O morphotype. Similarly, proteins involved in the carbohydrate metabolism are

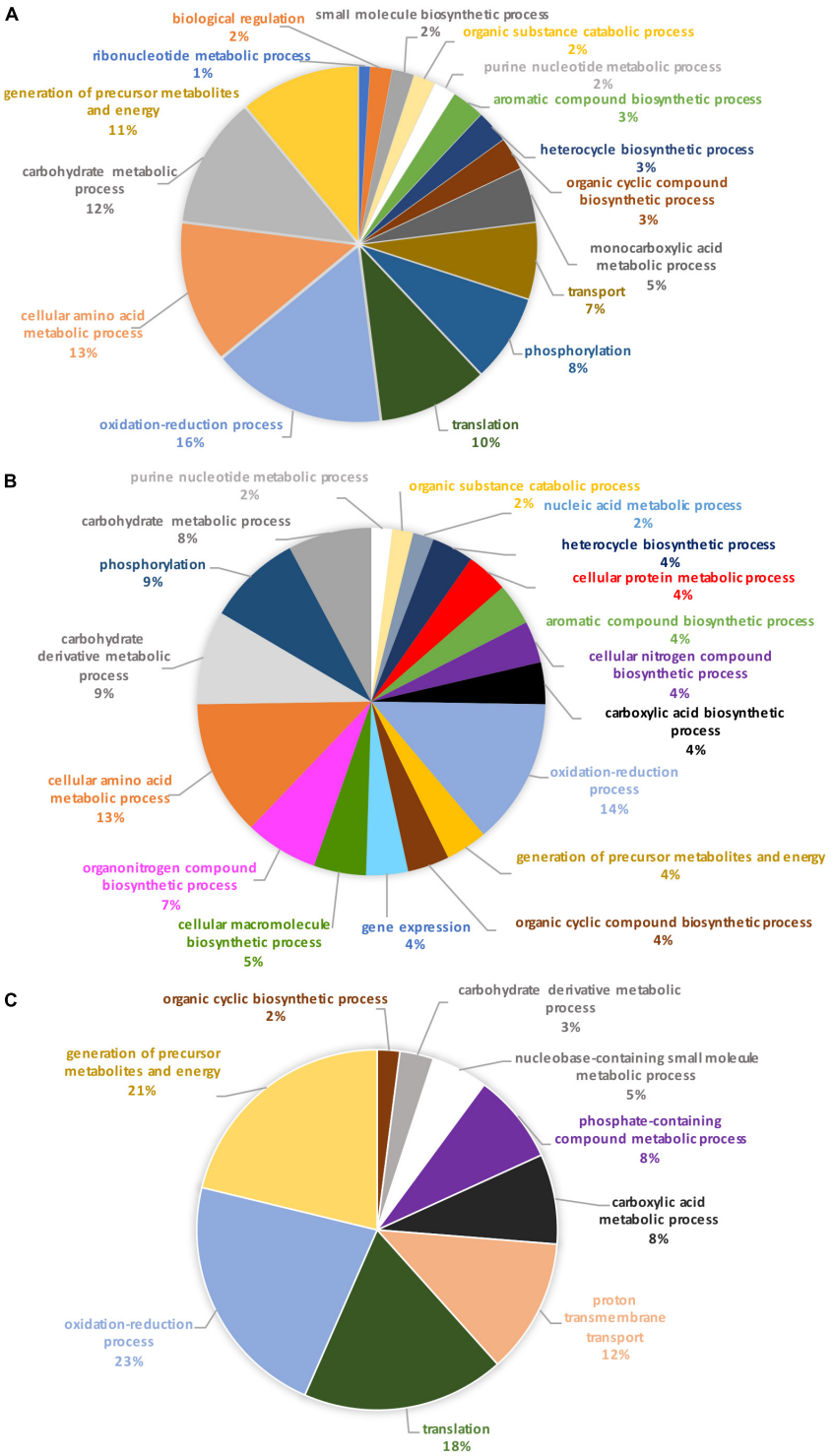


FIGURE 3 | Pie charts representing the biological processes, which are alliterated in the OF pairwise comparison. **(A)** Overall biological processes associated with the overall differentially expressed proteins in the oval morphotype when compared to the fusiform one. **(B)** Biological processes which are associated with the up-regulated proteins. **(C)** Biological processes associated with down-regulated proteins.

either up- (9%) or down-regulated (3%). These results suggest a fine-tune regulation of proteins involved in these specific biological processes (**Figures 3B,C**).

As previously described in the OF volcano plot (**Figure 2**), an abundance of 61 proteins changed at least twofold with *p*-values lower than 0.05. Among the 23 up-regulated proteins, 13

TABLE 2 | Biological process in which oval morphotype up-regulated proteins are involved.

SeqName	ProteinName	Description	Biological function
Phatr3_Jdraft1820	B7S4B2	Alcohol dehydrogenase	Glycolysis
Phatr3_J18911	B7FTW1	Aspartate–ammonia ligase	Acid amino synthesis
Phatr3_EG02188	B7FPT2	Protein *S*-acyltransferase	Protein degradation
Phatr3_J34976	B7FX80	Glutathione *S*-transferase mu 3	Cellular response to chemical stimulus
Phatr3_J43466	B7FSB5	Cytochrome P450	Oxidation-reduction process
Phatr3_J45046	B7FWA5	CBS domain-containing protein	Metabolic and cellular process
Phatr3_J45621	B7FYB0	NO-inducible flavohemoprotein	Response to nitrosative stress
Phatr3_J35939	B7FZX8	SDR family oxidoreductase	Oxidation-reduction process
Phatr3_J49119	B7G9J7	Alpha/beta hydrolase	Catalytic and hydrolase activity
Phatr3_J37667	B7G3E9	SDR family oxidoreductase	NA
Phatr3_J49937	B7GCD4	Predicted protein	NA
Phatr3_EG02230	B7G9U5	Predicted protein	NA
Phatr3_J46597	B7G1T2	Predicted protein	NA
Phatr3_J50914	B7FZJ4	Alkene reductase	Oxidation-reduction process
Phatr3_EG02330	B7G2D1	Predicted protein	NA
Phatr3_J33876	B7FU42	Predicted protein	NA
Phatr3_J15393	B7G884	NAD(P)H:quinone oxidoreductase, type IV	Oxidation-reduction process
Phatr3_J47823	B7G516	Predicted protein	NA
Phatr3_J44546	B7FUG6	Aldo/keto reductase	Oxidation-reduction process
Phatr3_J44092	B5Y5B5	Predicted protein	NA
Phatr3_Jdraft1693	B7S462	Predicted protein	NA
Phatr3_J12416	B7FZQ2	Hydroxylamine reductase	Oxidation-reduction process
Phatr3_J47840	B7G535	Predicted protein	NA

Only proteins with a log fold change greater than 2 are presented. NA: no biological process was associated with the Blast2GO software.

exhibits at least one known function and are mainly involved in the oxidation-reduction process like the cytochrome P450, the SDR family oxidoreductase, alkene reductase, and hydroxylamine reductase (**Table 2**). Ten proteins from this subset do not display either description or biological function (**Table 2**). Similarly, among the 38 down-regulated proteins, 23 of them exhibited a biological function in relation to cellular adhesion, transport, and catalytic activity (**Table 3**), whereas the functions of 15 proteins were not elucidated yet as they could not be related to any known biological function.

When comparing the pie chart representing the TF biological process pairwise comparison (**Figure 4**), we have shown that

TABLE 3 | Biological process in which oval morphotype down-regulated proteins are involved.

SeqName	ProteinName	Description	Biological function
Phatr3_EG02655	B7FXJ4	Fasciclin domain-containing protein	Cellular adhesion
Phatr3_J45403	B7FXK5	Predicted protein	NA
Phatr3_J44526	B7FUE7	Predicted protein	Carbonic anhydrase alpha enzyme
Phatr3_J50592	B7GEL9	Predicted protein	Cell wall/membrane/envelope biogenesis
Phatr3_J46400	B7G133	Predicted protein	Catalytic activity
Phatr3_J49202	B7G9T6	Predicted protein	NA
Phatr3_J49297	B7GA49	Predicted protein	NA
Phatr3_J45402	B7FXK4	Predicted protein	NA
Phatr3_J48730	B7G866	Predicted protein	NA
Phatr3_J48704	B7G7F7	Methyltransferase domain-containing protein	Methyltransferase activity
Phatr3_J49567	B7GB24	Predicted protein	DNA-binding transcription factor activity
Phatr3_J45464	B7FXS7	Predicted protein	NA
Phatr3_EG02527	B5Y460	V-type H(+)-translocating pyrophosphatase	Inorganic diphosphatase activity
Phatr3_EG02354	B7FVR9	Predicted protein	NA
Phatr3_J17519	B7FPK3	40S ribosomal protein IP6	Structural constituent of ribosome
Phatr3_J48383	B7G6Y2	Predicted protein	NA
Phatr3_J54686	B7G2A6	Predicted protein	Catalytic activity
Phatr3_EG02167	B7FQE7	Predicted protein	NA
Phatr3_EG02265	B7G6X2	Predicted protein	3′,5′-cyclic-nucleotide phosphodiesterase activity
Phatr3_J46046	B7FZV5	Predicted protein	NA
Phatr3_J49296	B7GA48	Predicted protein	NA
Phatr3_J41518	B7GEF5	Predicted protein	Methyltransferase activity
Phatr3_J50019	B7GCM8	Predicted protein	NA
Phatr3_J47667	B7G4H1	Predicted protein	Sodium-dependent phosphate transmembrane transporter activity
Phatr3_J47412	B7G3A5	Predicted protein	Nucleic acid binding
Phatr3_J48827	B7G8I4	Predicted protein	NA
Phatr3_J46547	B7G1L3	Predicted protein	Serine-type endopeptidase activity
Phatr3_J40158	B7GAM6	Predicted protein	Nucleic acid binding
Phatr3_J54642	B5Y3R0	Transitional endoplasmic reticulum ATPase	ATP binding
Phatr3_J12989	B7G1S8	NarL family transcriptional regulator	Catalytic activity
Phatr3_J49215	B7G9V1	Predicted protein	NA
Phatr3_J52619	B7GBV6	Purine permease	Transmembrane transporter activity

(Continued)

TABLE 3 | (Continued)

SeqName	ProteinName	Description	Biological function
PHATRDRAFT_55198	B7GE39	STT3 subunit-like protein	Protein amino acid glycosylation
PHATR_10209	B5Y4T4	Coatomer subunit gamma	Vesicle transport intracellular transport
PHATRDRAFT_45808	B7FYR4	Predicted protein	NA

Only proteins with a log-fold changed lower than 2 are presented.
NA, no biological process was associated with the Blast2GO software.

the oxidation-reduction biological process represents 27%, the glycolytic process 13%, the proteolysis 11%, the protein folding 10%, thus covering above 61% of the differentially produced proteins. In addition, proteins involved in the biosynthetic process of the branched amino acids, in the glucose metabolic process, in photosynthesis, and in light-harvesting accounted for 6% each. Other GO biological processes like transmembrane and intracellular transport, gene expression, cellular macromolecule biosynthetic process, and cellular component biogenesis were also found.

Specific Metabolic Pathways Are Involved in the Oval and Triradiate Morphotypes

In order to identify the proteins, which can be specifically assigned to either the oval or triradiate morphotypes, intersection analysis was performed and presented by the Venn diagram in **Figure 5**. This figure demonstrates that 32 proteins were identified at the intersection of the OF and TF pairwise comparisons, thus being specific to fusiform morphotype (**Supplementary Data 1**). A large portion of these 32 proteins identified is related to the oxido-reduction process (31%), followed by protein involved in proteolysis (16%) and protein folding (6%). In total, these three categories represented up to 53% of the biological processes. Other categories such as glycolytic and glucose metabolic processes represent 6% of the biological processes identified in the F morphotype (**Supplementary Figure 2** and **Supplementary Data 1**). Sixteen proteins were identified to be specific to the T morphotype (**Figure 5**). Among these T-specific proteins, some are belonging to protein folding, photosynthesis/light-harvesting, oxido-reduction process, proteolysis (**Supplementary Figure 3** and **Supplementary Data 2**). As far as the Oval morphotype is concerned, 648 proteins were identified to be specifically produced in the O morphotype in the intersection analysis (**Figure 5**). Among these 648 specific proteins, Blast2GO analysis was able to annotate and attribute GO biological processes to 544 proteins. Cellular amino acid metabolic process (12%), organophosphate metabolic process (11%), oxido-reduction process (14%), generation of precursor metabolites and energy (10%), phosphate-containing compound (11%), carbohydrate derivative process (10%) and translation (8%) are the most represented GO biological processes in O morphotype (**Figure 6A**). More precisely, our results show that phosphorylation (8%), cellular amino acid metabolic process (14%), carbohydrate (8%), and organonitrogen compound biosynthetic process (7%) are specifically up-regulated in the O cells (**Figure 6B**) whereas processes such as translation (15%), generation of precursor metabolites and energy (18%), organic substance transport (12%) and proton transmembrane transport (10%) are down-regulated (**Figure 6C**). To get information on differentially produced proteins involved in the specific biological pathway, analysis of the protein families was carried out with PANTHER. This analysis revealed that some proteins families are significantly overrepresented in the O cells compared to the other morphotypes. Many of the GO terms that were overrepresented were related to the purine nucleoside metabolism with a fold enrichment close to 6, to the ATP metabolic process (fold enrichment 5.5), and to the cellular amino acid biosynthetic process (fold enrichment 4.5) (**Figure 7**). The overrepresentation of these terms could indicate an expanded network that synthesizes specific metabolites in O cells.

To confirm previous results obtained with the label-free proteomic approach, we decided to perform quantitative proteomic analysis using iTRAQ® (for isobaric tags for relative and absolute quantification) labeling. In this experiment, the intracellular proteins extracted from the O, the T, and the F morphotypes were tagged respectively with iTRAQ® reagents 115, 116, and 117. Differential protein production was defined as an iTRAQ® ratio between the O or T morphotype using the F morphotype as a reference and ratios with a fold change higher than 1.2 were considered (**Supplementary Data 3**). The outcome of the iTRAQ® analysis allowed identifying 330 proteins, which are strictly identical to the subset of the 696 proteins identified with the label-free approach, thus corresponding to above 44% of the proteome (**Figure 8**). Only 2% of contradictory results were observed on a total of 330 common proteins (**Figure 8**). Results showing the differentially expressed proteins with the iTRAQ® methods and label-free analysis are presented in **Supplementary Figure 3**. Overall, the iTRAQ® results are consistent with those of the label-free analyses regarding the quantification of the differentially produced proteins. These results reveal that similarly to the label-free analysis, the proteasome, nucleoside metabolic process, photosynthesis, carbohydrate, and cellular amino acid biosynthesis were overrepresented highlighting a good correlation and the robustness of the data resulting from the label-free and iTRAQ® analyses. However, the label-free approach allowed us to identify more proteins, so we decided to use this method for the analysis of the secretomes.

Differences in Extracellular Secreted Protein Productions Exist Between the Three Pt3 *Phaeodactylum tricornutum* Morphotypes

Proteins that are secreted in the culture media by the diatom cells constitute the secretome. Such proteins may play important roles in cell migration, cell signaling, defense, and communication. These proteins can also be degraded into amino acids and serve as a source of nitrogen and/or carbon.

In order to further characterize the *P. tricornutum* morphotypes, we decided to identify the secretome of the

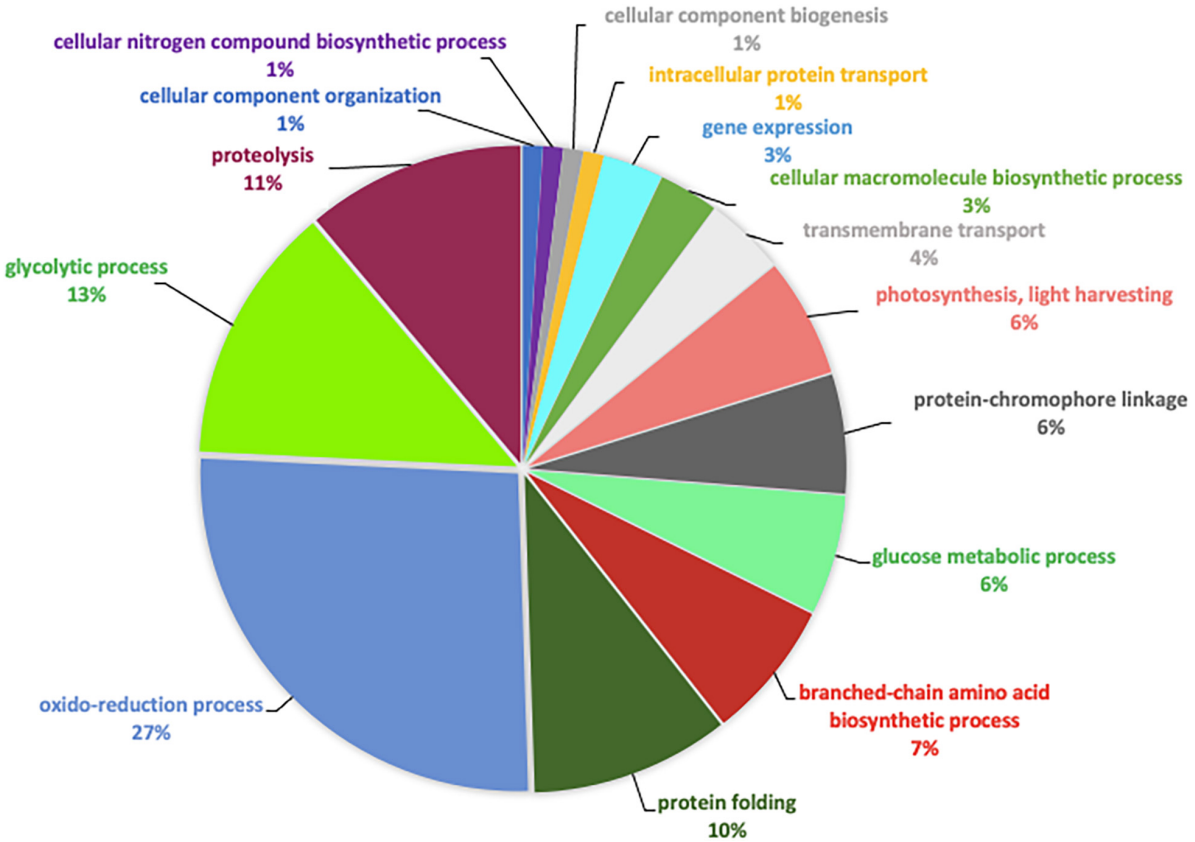


FIGURE 4 | Pie chart representing the biological processes, which are alliterated in the TF pairwise comparison.

previous enriched cultures. Overall, 949 secreted proteins were identified in the culture media of the three morphotypes with the label-free analysis. As previously described, Blast2GO software was used to attribute GO annotation and determine the biological functions in which these proteins are involved. From this analysis, above 87.5% of the secreted proteins have been annotated with GO terms thus corresponding to a total of 831 proteins (375 proteins for O, 207 for T, and 251 for F secreted proteins). Repartition of the identified proteins based on their biological process was accomplished. Pie chart representing the F biological process (**Figure 9A**) shows oxidation-reduction (19%), cellular amino acid (12%), carbohydrate (14%), phosphorylation, and proteolysis (11%) processes represent overall 56% of the differentially produced proteins. The response to stimulus and regulation of cellular process represent 4 and 5% while purine nucleotide metabolic pathway and cellular nitrogen biosynthetic pathway represent 2% of the biological process. Concerning the proteins which are secreted in the culture media of the T cells, it appears that the biological processes in which they are involved are relatively similar to the ones of F secreted proteins (**Figure 9B**). Proteins involved in oxidation-reduction processes represent up to 16%, followed by protein involved in cellular amino acids (14%), carbohydrates (10%), phosphorylation (9%), and proteolysis (11%), thus representing overall 60% of the differentially produced proteins. Interestingly, the biological process associated with macromolecule modification (1%), aromatic compound biosynthetic pathway (1%) appears to be specifically present in the triradiate secreted proteins. As far as the secreted proteins from the O morphotype are concerned, biological processes such as proteolysis (21%) and cellular

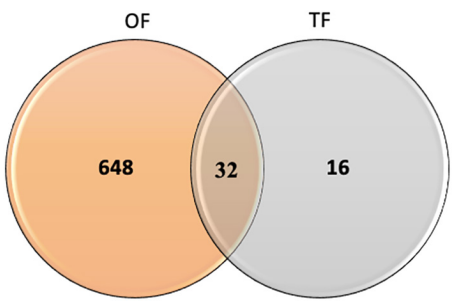


FIGURE 5 | Venn diagram displaying the specific features of proteins identified in the Oval vs. Fusiform (OF) and Triradiate vs. Fusiform (TF) pairwise comparisons. Thirty-two proteins were overlapping between the OF and TF comparisons. Six hundred and forty-eight and 16 proteins were specifically expressed in oval and triradiate morphotypes respectively. Venn diagram analysis was achieved using the 680 proteins subset identified in the OF pairwise comparison and the 48 proteins identified for the TF pairwise comparison (p-value ≤ 0.05).

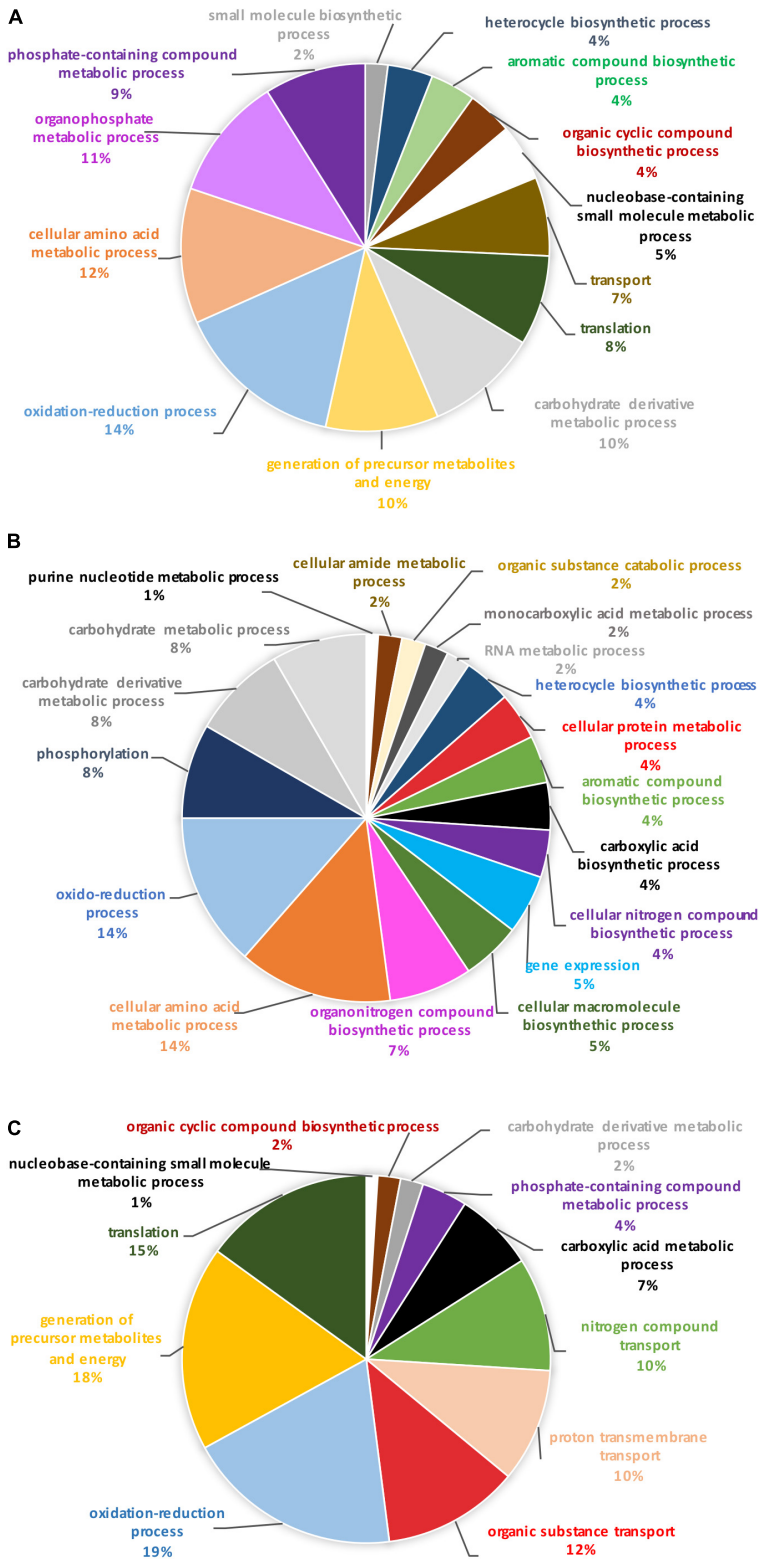


FIGURE 6 | Pie chart representing the biological process in which the 648 proteins, unique to the oval morphotype are involved. **(A)** Overall biological processes associated with the overall specific expressed proteins in the oval morphotype. **(B)** Biological processes which are associated with the up-regulated proteins. **(C)** Biological processes associated with down-regulated proteins.

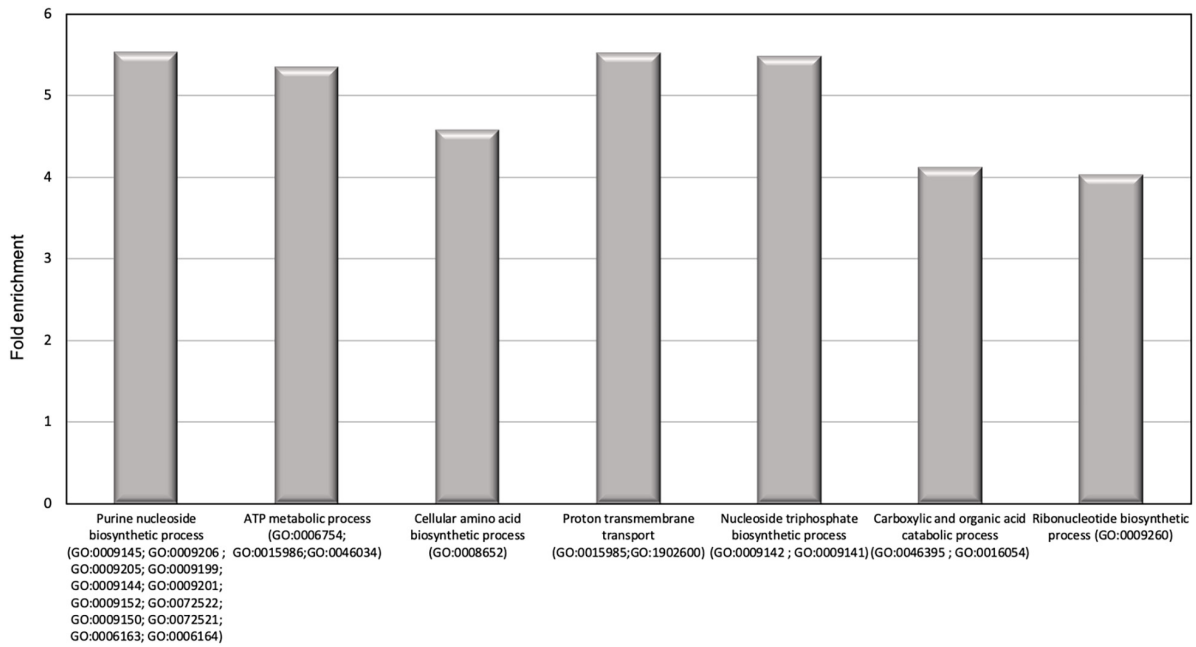


FIGURE 7 | Results of the overrepresentation test (Panther) of the 648 proteins specifically expressed in oval morphotype. Only proteins families with a fold enrichment upper than 4 are represented.

nitrogen compound biosynthetic pathway (4%) are increased in comparison to the two other morphotypes (**Figure 9C**).

To go further in the characterization of *P. tricornutum* secretomes, an intersection analysis was run in order to identify the proteins, which can be specifically attributed to either the oval, fusiform, or triradiate secretomes. The resulting Venn diagram presented in **Figure 10** demonstrated that among the identified secreted proteins, 22, 10, and 163 secreted proteins were specific to the F, T, and O morphotypes, respectively (**Supplementary Data 4**). Interestingly, this result clearly confirmed previous observation (Ovide et al., 2018; Song et al., 2020; Galas et al., 2021) that the O cells are secreted more proteins than the two others morphotypes. Again, the F and T secretomes appeared to be very similar (**Figure 10**). Analysis of the protein families by PANTHER highlighted that proteasome catabolism (fold enrichment 11), protein catabolism (fold enrichment 7), and acid catabolic process (fold enrichment 6) are significantly overrepresented in the medium of O cells compared to the other morphotypes (**Figure 11**). No specific enrichment was observed for secreted proteins of F and T cells probably due to the fact that the number of specific proteins for both morphotypes was relatively low (22 and 10, respectively).

The Secretomes of the Three Morphotypes Are Particularly Enriched in Glycoproteins

In order to confirm the proteins identification in the secretomes of the different morphotypes, we look for the presence of a putative signal peptide and *N*-glycosylation consensus sites that are characteristics of proteins intended to travel through the secretory pathway and to be secreted in the culture medium. The use of the SignalP online tool allows us to predict the presence of a signal peptide for more than 47.5% of proteins in the F secretome, almost 40% of proteins in the T one, and 50% in the O one. It is to note that only predictions of signal peptides with a probability greater than 0.5 were considered (**Supplementary Data 5**, **excel sheet 1**). Indications regarding the position of the signal peptide are also available (**Supplementary Data 5**, **excel sheet 1**). For

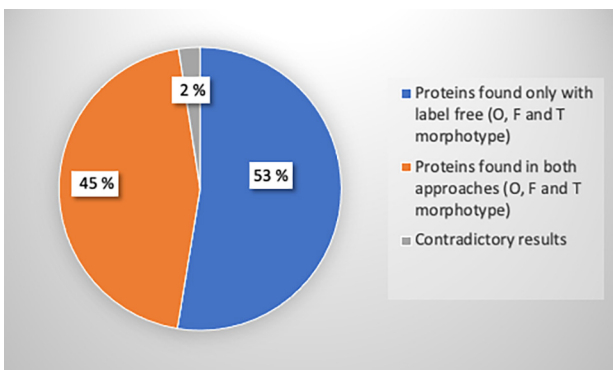


FIGURE 8 | Pie chart representing the number of proteins identified with iTRAQ®, label-free or both methods, and the proportion of contradictory results. A total of 696 proteins were identified with label-free analysis (p-value ≤ 0.05): 648 proteins for oval, 32 proteins for fusiform, and 16 proteins for triradiate morphotypes. A total of 330 proteins were identified with iTRAQ® analysis. Among these proteins, 17 proteins representing 2% of iTRAQ® identified proteins were differentially regulated between both approaches.

comparison, prediction of the signal peptide of the intracellular proteins identified for each morphotype has been looked for, showing an average of 26.5% predicted signal peptide in the overall intracellular proteins. With regards to the *N*-glycosylation predictions that were run using the NetNGlyc tool, 67% of the F secreted proteins, 64.5% of the T secreted proteins, and finally, 71.3% of the O secreted proteins were predicted to bear *N*-glycosylation sites, respectively. This represents an average of 2.67 *N*-glycosylation sites per secreted protein independently of the morphotypes (**Supplementary Data 5**, **excel sheet 1**).

Analysis of the Putative Evolutive Origin of the Secreted Proteins

In order to determine the origin of the secreted proteins that are secreted by the different morphotypes of *P. tricornutum* Pt3 strain, we compared the dataset from this study to the one reported in Dorrell et al., 2021 (**Supplementary Data 5**, **excel sheets 2–5**). The comparison allows us to determine that among the 274 secreted proteins that were identified in the F morphotype, only 22 were also present in the Dorrell et al. (2021) dataset. Such subset is enriched mainly in proteins originating from bacteria (8) representing 36.4%, alveolates (6) representing 27.2%, and dinoflagellates (4) accounting for 18.2% (**Supplementary Data 5**, **excel sheets 2, 5**). A similar trend was observed for 23 out of 230 proteins identified in the T morphotype. Among these proteins, 7 originated from bacteria (30%), 6 from alveolates (26%), and finally 5 from dinoflagellates (21.7%) (**Supplementary Data 5**, **excel sheets 3, 5**). As far as the secreted proteins from the O morphotype are concerned, only 13 out of 384 were common with previous data from Dorrell et al. (2021). Among the 13 proteins, 5 are originated from bacteria, thus representing 38.4% and 3 originated from dinoflagellates (23%), the others coming from alveolates (2), haptophytes (2), and Rhizaria (1) lineages (**Supplementary Data 5**, **excel sheets 4, 5**). Finally, the list of 13 proteins from the overall O secretome was crossed with the 163 proteins that are specific to the O morphotype. It came out from this comparison that only 3 were specific to the O morphotype: the protein B7G0Q2 (Phatr3_J20677) and the protein B7FZM8 (Phatr3_J35518) originated from bacteria whereas the protein B5Y4E4 (Phatr3_J33120) derived from Rhizaria. The percentage of secreted proteins originating from bacteria displays above 35% of the secretome independently of the morphotype. This number is below the 49% observed in the overall genome of *P. tricornutum* for which bacteria were reported to be responsible for nearly half of the gene transfers (Dorrell et al., 2021).

Comparative Analysis of the Proteome Versus Transcriptome

In order to evaluate the relationship between previous transcriptomic data reported on the Pt3 morphotypes and the current proteomic data, we analyzed and compared the proteome of the O cells to the transcriptomes reporting in Ovide et al. (2018) that compared O vs. F (OF) and O vs. T (OT). Among the 648 intracellular proteins that were specifically identified in the O proteomic data, 35.65% were also differentially expressed in the transcriptome data comparing OT and OF. A similar comparison was made for the 163 secreted proteins that have been identified in the secretome of the O cells and above 32.7% were also differentially expressed in the transcriptome dataset (**Supplementary Data 6**).

DISCUSSION

The diatom *P. tricornutum* exists under three distinct major morphotypes, fusiform, triradiate, and oval cells that can be observed mostly depending on the environmental conditions (Tesson et al., 2009). The fusiform morphotype is considered the most common one (De Martino et al., 2007, 2011; Ovide et al., 2018). As a consequence, mainly research performed on *P. tricornutum* fusiform morphotype is widely available (Xie et al., 2015; Bai et al., 2016; Longworth et al., 2016). However, a transcriptome-wide characterization of the three morphotypes originating from the Pt3 strain has recently been performed revealing differences in transcriptomic regulation between morphotypes (Ovide et al., 2018). Moreover, recent biochemical and imaging analyses revealed that the oval morphotype is synthesizing more proteins (Song et al., 2020) and is secreting proteins more rapidly as compared to the other morphotypes (Galas et al., 2021). However, so far, no proteomic studies of the three morphotypes have been carried out. In the present work, a total of 728 intracellular proteins of *P. tricornutum* were identified. Pairwise comparisons showed that fusiform and triradiate morphotypes were very similar as only the expression of 48 intracellular proteins were differentially regulated between these two morphotypes whereas 680 intracellular proteins were differentially regulated when comparing the oval morphotype versus the fusiform one. This result is consistent with previous reports that reported fusiform and triradiate morphotypes as similar morphotypes (less than 1% of the transcriptomes were significantly differentially expressed, Ovide et al., 2018) whereas the oval cells were described to be more active metabolically (Bartual et al., 2008; Ovide et al., 2018; Song et al., 2020). Interestingly, in our study, 648 intracellular proteins appeared to be oval specific confirming that the oval morphotype is more metabolically active.

Biological processes such as purine (ribo)nucleoside metabolic pathway, carbohydrate, and cellular amino acid biosynthesis are overrepresented in the oval cells while representing only 2% of the overall intracellular proteins. These findings indicated that the oval morphotype promotes nucleotide, carbohydrate, and amino acid biosynthesis. Such an increase has been previously reported by Ovide et al. (2018) where 68% of genes involved in the primary metabolism pathway (glycolysis, nucleotide, etc.) were increased in the OF pairwise comparison. The increased production of proteins responsible for various biosynthetic processes suggests that the oval cells are extremely responsive to their environment. It has been hypothesized that the oval cells are more resistant to stresses as this morphotype seems to be favored under extreme conditions such as low temperature, the presence of bacteria, or low salinity (De Martino et al., 2011; Buhmann et al., 2016). In our study, upregulation (>2-fold) of proteins

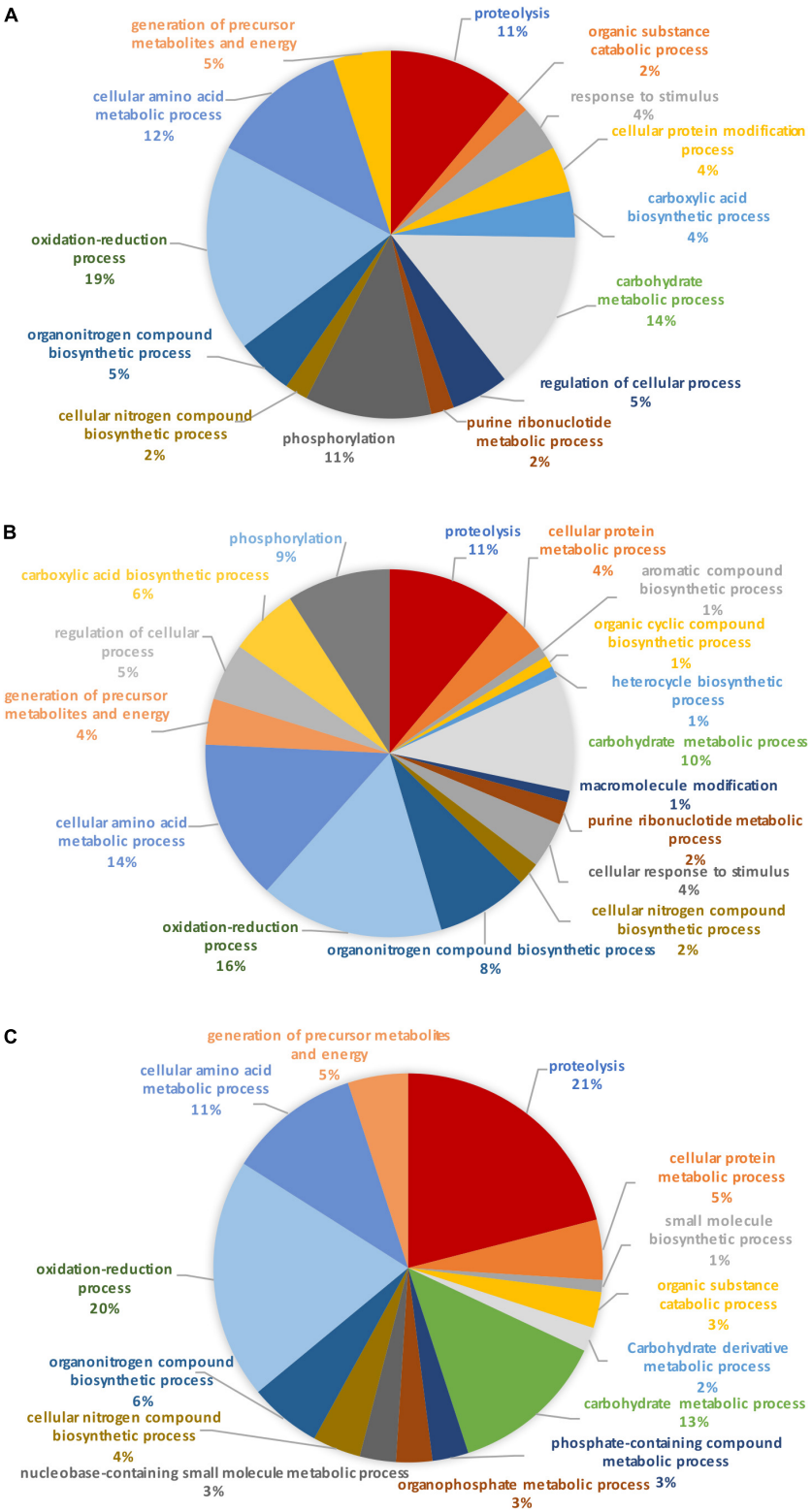


FIGURE 9 | Pie charts representing the biological process in which **(A)** fusiform, **(B)** triradiate, and **(C)** oval morphotype secreted proteins are involved. Pie charts are representative of 5 biological replicates. **(A)** Biological processes of the 275 proteins secreted in fusiform morphotype. **(B)** Biological processes of the 231 proteins secreted in triradiate morphotype. **(C)** Biological processes of the 385 proteins secreted in oval morphotype.

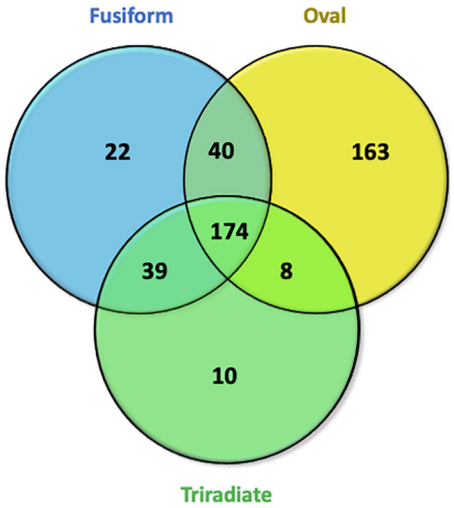


FIGURE 10 | Venn diagrams displaying overlaps of secreted proteins identified in Oval, Fusiform, and Triradiate cells. In the total proteins, 174 secreted proteins were identified in the three groups. Thirty-nine proteins were present in fusiform and triradiate cells, 40 proteins were present in fusiform and oval cells, 8 proteins were present in triradiate and oval cells. 22, 10, and 163 appear to be specific to fusiform, triradiate, and oval cells, respectively. Venn diagrams are representative of 5 biological replicates for each morphotype.

such as B7FX80, a glutathione S-transferase mu 3 implicated in the cellular response to stimulus, and B7FYB0, a NO-inducible flavohemoprotein involved in response to nitrosative stress, as well as proteins involved in the oxidative process, was consistent with this hypothesis and reflects the specific regulation of stress-responsive proteins. However, we should keep in mind that oval cells were cultured in 10 % seawater in order to reach an oval cells enrichment of 98% but also because it was not possible to maintain an enriched culture in oval morphotype using 100% seawater media (Ovide et al., 2018). The reverse was not possible with fusiform and triradiate cells in 10% seawater as most cells switched into oval cells when maintained in such culture conditions. As a consequence, we hypothesize that the oval cells present a metabolism specifically adapted to environment variables. Oval cells may contain specific proteins involved in repair mechanisms in contrast to the fusiform and triradiate cells allowing the oval cells to survive when conditions become unfavorable. The presence of such proteins in oval cells may explain the interconversion of fusiform or triradiate morphotype into this more resistant morphotype.

This agrees with the fact that diatoms have developed defense mechanisms to overcome unfriendly environments (Falciatore et al., 2020). In this study, we noticed the presence of proteins such as B7FPT2 identified as a palmitoyltransferase, B7FUR0, and B7FYUX1 involved in terpenoid biosynthesis in the oval cells. This may suggest the production of defense molecules in this morphotype. Similarly, earlier studies demonstrated that *P. tricornutum* was able to synthesized molecules known to exert antimicrobial activity such as terpenoids (α-1, 8-cineole, α-pinene, limonene; Prestegard et al., 2015), palmitoleic acid (Desbois et al., 2009), or reactive oxygen species (Buhmann et al., 2016) in higher amounts in oval cells. Altogether, it appears likely that oval cells synthesize molecules specifically involved in defense pathways against predators or pathogen organisms. Interestingly, the two glyceraldehyde-3-phosphate dehydrogenases (B7G5Q1; B7G6K6), the oxygen-evolving enhancer protein B7F296 and the phosphoribulokinase B5Y5F0 of which abundances were reported to increase in fusiform cells after 4 days of darkness (Bai et al., 2016), are more abundant in Pt3 oval cells in comparison to Pt3 fusiform cells. As a consequence, we hypothesize that oval cells naturally synthesized more proteins related to dark conditions compared to fusiform cells. To challenge this hypothesis, we compared the list of genes encoding proteins specific to the oval cells to the 104 genes with robust diel oscillating expression reported in Annunziata et al. (2019). The expression of the 104 genes was studied using a 16h:8h light: dark cycle as the one that was used in this study. However, once should keep in mind that the media used in the two studies to grow the cells are not exactly the same (f/2 Guillard media in Annunziata et al., 2019 and Conway in the present study) and that the harvest of the diatom cells in this study were done few hours after illumination. From that comparison, only 11 were found in the intracellular proteins of the oval cells, none of them being photoreceptors. Among the 11 candidates, 7 are implicated in the metabolism (Pds1, Gapc, Psy1, Gsat, Zep3, CaThioredoxin, and Zds), 2 belong to the cell cycle (FtsZ and Pcna) and the two last ones were transcription factors (HSF4.2j and HSF4.7a). This might also be correlated with the fact that fusiform cells have a planktonic lifestyle considering that oval cells are preferentially benthic morphotypes (De Martino et al., 2007; Stanley and Callow, 2007; Willis et al., 2013). Therefore, the difference in lifestyle exposed the morphotypes to light variation with the fusiform cells benefiting from a higher exposure. Moreover, it is to note that most of these proteins are involved in the carotenoid biosynthesis pathway (Pds, Psy, Zep, Zds) (Bertrand, 2010; Scarsini et al., 2020) and in photosynthesis (Gsat; Schoefs and Bertrand, 2005).

When comparing previous transcriptomic data (Ovide et al., 2018) and the proteomic datasets described here, some discrepancies and only 1/3 of positive correlation were observed. Such differences could be related to a fine-tuning gene expression that can be due to allele-specific expression (ASE). Such phenomena have been recently described in *P. tricornutum* (Hoguin et al., 2021). Such ASE genes were enriched in genes involved in catabolism processes including proteasome subunits proteins, and autophagy, in intracellular protein transport, exocytosis, and endocytosis (Hoguin et al., 2021) that correspond to pathways identified in this study. Moreover, it would also be interesting to study gene alternative splicing as it was reported previously that extensive alternative splicing was involved in the regulation of gene expression in response to nutrient starvation, suggesting that *P. tricornutum* used it to cope with environmental changes (Rastogi et al., 2018).

However, it should not be forgotten that in our study GO terms were attributed for only 608 proteins from a total of 728 intracellular proteins. This means that even 12 years after the genome of *P. tricornutum* was sequenced (Bowler et al.,

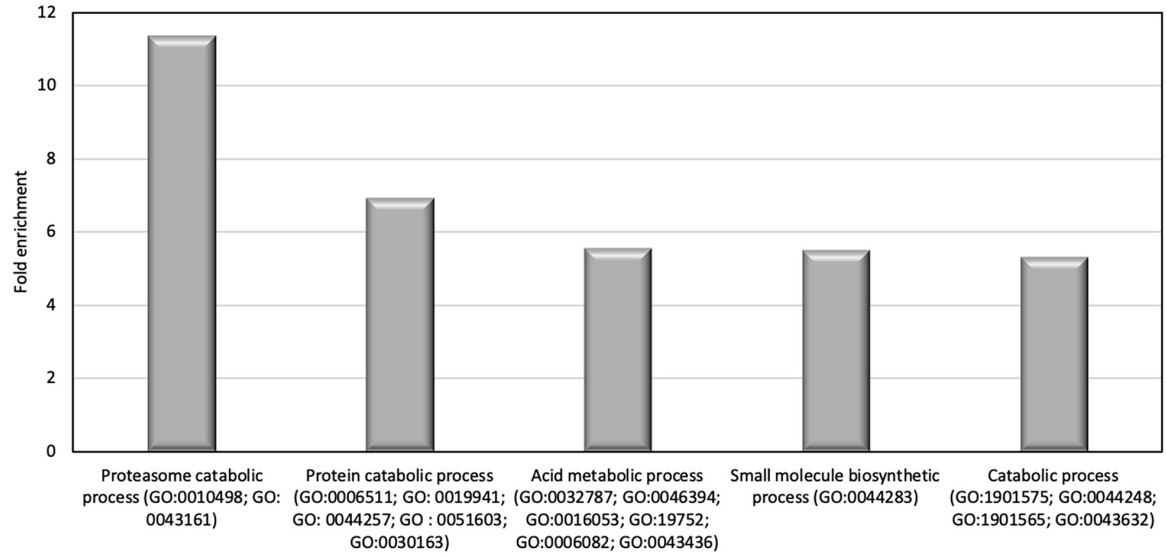


FIGURE 11 | Results of the overrepresentation test (Panther) of the 163 proteins found in the oval cell culture medium. Only proteins families with a fold enrichment upper than 4 are represented.

2008), there is at least 16 % of predicted proteins of which the biological function remains unanswered. In the future, a lot of work remains to be performed in order to identify the physiological role of those proteins in *P. tricornutum*. Moreover, it would be interesting to determine how many of the differentially accumulated proteins are encoded by genes that show SNPs. Indeed, Rastogi et al. (2020) recently revealed the global genetic polymorphism, structure, and functional diversity of ten accessions strains of *P. tricornutum* showing high differences in SNPs between ecotypes. Therefore, it would be interesting in future studies to analyze if proteins specific to the oval morphotype are more likely to contain SNPs. In this study, it was chosen to work with enriched cultures of oval, triradiate, and fusiform morphotypes issued from the same Pt3 strain and compared their specific proteomes and secretomes. It would be interesting in future studies to compare the proteome of Pt3 that is naturally rich in oval cells, one of the Pt1 (naturally rich in fusiform cells), and one of the Pt8 (rich in triradiate) and evaluate whether specific proteome signature can be identified for each morphotype regardless of the original strain.

To go further in the characterization of the *P. tricornutum* morphotype, we also expanded the analysis to the identification of the secreted proteins of each morphotype. It is to note that during the preparation of the samples, the culture medium was desalted and concentrated using a 3K MWCO Concentrator, thus, removing all peptides below 3 kDa. It will be interesting in a future study to characterize such a population of peptides that could be interesting small bioactive peptides differentially expressed between morphotypes.

As for the intracellular proteomic analysis, our results show that the number of secreted proteins was higher in the oval cells with a total of 385 proteins whereas in fusiform and triradiate cells several 275 and 231 secreted proteins were totalized, respectively. Among those proteins, 40 to 50% possess a predicted signal peptide with a probability greater than 0.5 while only 13.3% (1,629 proteins out of 12,233) to 15% (1,831 proteins out of 12,179) were reported in the entire proteomes of *P. tricornutum* in the study of Ait-Mohamed et al. (2020) and Rastogi et al. (2020), respectively. Moreover, in this study, as expected, the intracellular proteins present only 26.5% of the predicted signal peptides on average. The Venn diagram identified 163 proteins that were specific for the oval cells whereas only 22 and 10 secreted proteins were specific for fusiform and triradiate morphotypes, respectively. This result, in agreement with previous reports, demonstrates that in the oval morphotype, more proteins are secreted (Ovide et al., 2018; Song et al., 2020). This higher number of secreted proteins in the oval cells can be related to the excretion of exopolymeric substances that favor cell adhesion and colony formation specific to the oval morphotype (Stanley and Callow, 2007; Willis et al., 2013). The difference in adhesion of *P. tricornutum* cells was reported to be dependent on the composition of the various types of EPS produced (Stanley and Callow, 2007), which can be explained by differences in the proportion of monosaccharides, chain terminal saccharide and the degree of sulphation. For example, salinity changes resulted in an increase of carbohydrate production, with enrichment of highly branched/substituted and terminal rhamnose, xylose, and fucose as well as *O*-methylated sugars, uronic acids, and sulfate (Abdullahi et al., 2006).

Recently, Erdene-Ochir et al. (2019) reported that the protein B7G4A0, also known as HASP1 protein, is the most abundant protein secreted into the culture medium of *P. tricornutum*. Interestingly, this finding is also supported by our results as we noted the presence of B7G4A0 in the culture media of the fusiform and triradiate cells. By contrast, we do not report the presence of this protein in the culture medium of the oval cells. In addition, our results also confirmed that the proteins B7FSH1 and B7G259 reported being the third and the

fourth most secreted proteins reported by Erdene-Ochir et al. (2019), are also detected in the culture media of the three morphotypes. In addition, the proteins B5Y3F2 and B7GBF3, which were reported to be abundant in the culture media of *P. tricornutum*, were only detected in the oval and triradiate culture media in our conditions. Such results are of particular interest as the HASP1 gene promoter for example has been used to improve the production of secreted recombinant proteins by *P. tricornutum* (Erdene-Ochir et al., 2019). In addition, our results highlight the importance of the morphotypes for the efficient expression and secretion of heterologous recombinant proteins when *P. tricornutum* is intended to be used as a cell biofactory (Butler et al., 2020).

CONCLUSION AND FUTURE PROSPECTS

To date, our knowledge of the specificity of each morphotype of the model diatom *P. tricornutum* is rather limited. Many questions remain regarding the physiological significance of *P. tricornutum* morphogenesis as well as its mechanism of regulation. Indeed, little attention has been paid to the morphotype in *P. tricornutum* (De Martino et al., 2011; Ovide et al., 2018; Song et al., 2020). The results of the present study demonstrate that the oval morphotype appears to be unique and present a specific metabolic network compared to the fusiform and triradiate morphotypes. Moreover, our results confirm that the oval cells are secreting more proteins in the culture medium as previously suggested by Song et al. (2020) and Galas et al. (2021). This characteristic property in oval cells offers new and highly attractive prospects for the development of biopharmaceuticals in the diatom cell factory. Thus, it can be envisioned that the production yield of biopharmaceuticals like monoclonal antibodies, which is currently limited to a few micrograms per liter in *P. tricornutum* (Hempel et al., 2011; Hempel and Maier, 2012; Vanier et al., 2015, 2018; Hempel et al., 2017) could be increased by expression in *P. tricornutum* oval cells. This could be attempted by transforming oval cells or switching transformed fusiform cells within oval cells as previously reported (Tesson et al., 2009). Taking into account the specificity of the oval morphotype could help in the near future to further optimize the production of biopharmaceuticals in the diatom cell factory.

On the other hand, the presence of molecules with antimicrobial properties (Desbois et al., 2009; Prestegard et al., 2015; Butler et al., 2020) and of proteins involved in the biosynthesis of terpenoids highlight new potentialities of application for *P. tricornutum* such as protector of plant immunity. Cell-specific fractions from *P. tricornutum*, such as fatty acids or volatile organic compounds, were reported to possess antimicrobial activities against bacteria (Desbois et al., 2008, 2009; Prestegard et al., 2015). In addition, *P. tricornutum* is able to synthesized α(1-8)-cineole, α-pinene, limonene and also β(1-3)-glucanases that are present in higher concentrations in the oval morphotype (Prestegard et al., 2015; Buhmann et al., 2016). Such molecules are reported to be involved in plant defense response against plant pathogens (Balasubramanian et al., 2012; Lackus et al., 2018). In this context, the presence of a higher number of specific proteins suggests the potential of the oval cells to be good candidates for finding new elicitors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: The analyzed data are presented within the different tables and figures of this manuscript. Raw proteomics data are available in the MassIVE database, under accession number MSV000086835 (ftp://massive.ucsd.edu/MSV000086835/).

AUTHOR CONTRIBUTIONS

MB planned and designed the research. CC, BG, PC, CB, and JH performed the experiments. CC, BG, M-LW-B, FT, PC, and MB collected and analyzed the data. MB and CC interpreted the data. MB, CC, and PC wrote the manuscript. All authors have read and agreed on the manuscript prior to the submission.

FUNDING

The authors are grateful to the Region Normandie *via* the SFR NORVEGE (Project named Licorne), the Institut Universitaire de France (IUF), and the University of Rouen Normandie for financial support. PISSARO platform was co-supported by European Union and Région Normandie. Europe gets involved in Normandie with "European Regional Development Fund (ERDF)." Work performed in the Glyco-MEV lab, University of Rouen Normandie was partly funded by the "European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 774078 (Pharma-Factory)." The Ph.D. work of CC was funded by "The region Normandie." The Ph.D. work of M-LW-B was funded by the "European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 774078 (Pharma-Factory)." The funding bodies were not involved in the design of the study, collection, and interpretation of data nor the writing of the manuscript.

ACKNOWLEDGMENTS

The authors are thankful to Pr. Patrice Lerouge, University of Rouen Normandie for critical reading of the manuscript and to Chris Bowler, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR8197, Inserm U1024 for sharing its *P. tricornutum* Pt3 strain in 2010.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.673113/full#supplementary-material

REFERENCES

Abdullahi, A. S., Underwood, G. J. C., and Gretz, M. R. (2006). Extracellular matrix assembly in diatoms (Bacillariophyceae). V. Environmental effects on polysaccharide synthesis in the model diatom, *Phaeodactylum tricornutum*. *J. Phycol.* 42, 363–378. doi: 10.1111/j.1529-8817.2006.00193.x

Ait-Mohamed, O., Novák Vanclová, A. M. G., Joli, N., Liang, Y., Zhao, X., Genovesio, A., et al. (2020). PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 11:590949. doi: 10.3389/fpls.2020.590949

Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008). Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci. U S A* 105, 10438–10443. doi: 10.1073/pnas.0711370105

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Annunziata, R., Ritter, A., Fortunato, A. E., Manzotti, A., Cheminant-Navarro, S., Agier, N., et al. (2019). bHLH-PAS protein RITMO1 regulates diel biological rhythms in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. U S A.* 116, 13137–13142. doi: 10.1073/pnas.1819660116

Bai, X., Song, H., Lavoie, M., Zhu, K., Su, Y., Ye, H., et al. (2016). Proteomic analyses bring new insights into the effect of a dark stress on lipid biosynthesis in *Phaeodactylum tricornutum*. *Sci. Rep.* 6:25494. doi: 10.1038/srep25494

Balasubramanian, V., Vashisht, D., Cletus, J., and Sakthivel, N. (2012). Plant β-1,3-glucanases: their biological functions and transgenic expression against phytopathogenic fungi. *Biotechnol. Lett.* 34, 1983–1990. doi: 10.1007/s10529-012-1012-6

Bartual, A., Gálvez, J. A., and Ojeda, F. (2008). Phenotypic response of the diatom *Phaeodactylum tricornutum* Bohlin to experimental changes in the inorganic carbon system. *Botanica Marina* 51:47. doi: 10.1515/BOT.2008.047

Bertrand, M. (2010). Carotenoid biosynthesis in diatoms. *Photosynth. Res.* 106, 89–102. doi: 10.1007/s11120-010-9589-x

Borowitzka, M. A., and Volcani, B. E. (1978). The polymorphic diatom *Phaeodactylum tricornutum*: ultrastructure of its morphotypes. *J. Phycol.* 14, 10–21.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Buhmann, M. T., Schulze, B., Förderer, A., Schleheck, D., and Kroth, P. G. (2016). Bacteria may induce the secretion of mucin-like proteins by the diatom *Phaeodactylum tricornutum*. *J. Phycol.* 52, 463–474. doi: 10.1111/jpy.12409

Butler, T., Kapoore, R. V., and Vaidyanathan, S. (2020). *Phaeodactylum tricornutum*: A Diatom Cell Factory. *Trends Biotechnol.* 38, 606–622. doi: 10.1016/j.tibtech.2019.12.023

De Martino, A., Bartual, A., Willis, A., Meichenin, A., Villazán, B., Maheswari, U., et al. (2011). Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in the Model Diatom Phaeodactylum tricornutum. *Protist* 162, 462–481. doi: 10.1016/j.protis.2011.02.002

De Martino, A., Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *J. Phycol.* 43, 992–1009. doi: 10.1111/j.1529-8817.2007.00384.x

Desbois, A. P., Lebl, T., Yan, L., and Smith, V. J. (2008). Isolation and structural characterisation of two antibacterial free fatty acids from the marine diatom, *Phaeodactylum tricornutum*. *Appl. Microb. Biotechnol.* 81, 755–764. doi: 10.1007/s00253-008-1714-9

Desbois, A. P., Mearns-Spragg, A., and Smith, V. J. (2009). A Fatty Acid from the Diatom *Phaeodactylum tricornutum* is Antibacterial Against Diverse Bacteria Including Multi-resistant Staphylococcus aureus (MRSA). *Mar. Biotechnol.* 11, 45–52. doi: 10.1007/s10126-008-9118-5

Dorrell, R. G., Villain, A., Perez-Lamarque, B., Audren, de Kerdrel, G., McCallum, G., et al. (2021). Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc. Natl. Acad. Sci. U S A.* 118:e2009974118. doi: 10.1073/pnas.2009974118

Erdene-Ochir, E., Shin, B.-K., Kwon, B., Jung, C., and Pan, C.-H. (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Sci. Rep.* 9:9941. doi: 10.1038/s41598-019-45786-9

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. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158

Fréret, M., Drouot, L., Obry, A., Ahmed-Lacheheb, S., Dauly, C., Adriouch, S., et al. (2013). Overexpression of MHC class I in muscle of lymphocyte-deficient mice causes a severe myopathy with induction of the unfolded protein response). *Am. J. Pathol.* 183, 893–904. doi: 10.1016/j.ajpath.2013.06.003

Galas, L., Burel, C., Schapman, D., Ropitaux, M., Bernard, S., Bénard, M., et al. (2021). Comparative Structural and Functional Analyses of the Fusiform, Oval, and Triradiate Morphotypes of *Phaeodactylum tricornutum* Pt3 Strain. *Front. Plant Sci.* 12:638181. doi: 10.3389/fpls.2021.638181

Gutenbrunner, S. A., Thalhamer, J., and Schmid, A.-M. M. (1994). Proteinaceous and immunochemical distinctions between the oval and fusiform morphotypes of *Phaeodactylum tricornutum* (Bacillariophyceae). *J. Phycol* 30, 129–136. doi: 10.1111/j.0022-3646.1994.00129.x

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011). Algae as Protein Factories: Expression of a Human Antibody and the Respective Antigen in the Diatom *Phaeodactylum tricornutum*. *PLoS One* 6:7. doi: 10.1371/journal.pone.0028424

Hempel, F., and Maier, U. G. (2012). An engineered diatom acting like a plasma cell secreting human IgG antibodies with high efficiency. *Microb. Cell Fact.* 11:126. doi: 10.1186/1475-2859-11-126

Hempel, F., Maurer, M., Brockmann, B., Mayer, C., Biedenkopf, N., Kelterbaum, A., et al. (2017). From hybridomas to a robust microalgal-based production platform: molecular design of a diatom secreting monoclonal antibodies directed against the Marburg virus nucleoprotein. *Microb. Cell Fact.* 16:131. doi: 10.1186/s12934-017-0745-2

Heydarizadeh, P., Marchand, J., Chenais, B., Sabzalian, M. R., Zahedi, M., Moreau, B., et al. (2014). Functional investigations in diatoms need more than a transcriptomic approach. *Diatom Res.* 29, 75–89. doi: 10.1080/0269249X.2014.883727

Hoguin, A., Rastogi, A., and Bowler, C. (2021). Genome-wide analysis of allele-specific expression of genes in the model diatom *Phaeodactylum tricornutum*. *Sci. Rep.* 11:2954. doi: 10.1038/s41598-021-82529-1

Johansen, J. R. (1991). Morphological variability and cell wall composition of *Phaeodactylum tricornutum* (Bacillariophyceae). *Great Basin Naturalist.* 51, 310–315. doi: 10.2307/41712676

Lackus, N. D., Lackner, S., Gershenzon, J., Unsicker, S. B., and Köllner, T. G. (2018). The occurrence and formation of monoterpenes in herbivore-damaged poplar roots. *Sci. Rep.* 8 :17936. doi: 10.1038/s41598-018-36302-6

Lewin, J. C., Lewin, R. A., and Philpott, D. E. (1958). Observations on *Phaeodactylum tricornutum*. *J. Gen. Microb.* 18, 418–426.

Longworth, J., Wu, D., Huete-Ortega, M., Wright, P. C., and Vaidyanathan, S. (2016). Proteome response of *Phaeodactylum tricornutum*, during lipid accumulation induced by nitrogen depletion. *Algal Res.* 18, 213–224. doi: 10.1016/j.algal.2016.06.015

Mi, H., Muruganujan, A., and Thomas, P. D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 41, D377–D386. doi: 10.1093/nar/gks1118

Ovide, C., Kiefer-Meyer, M.-C., Bérard, C., Vergne, N., Lecroq, T., Plasson, C., et al. (2018). Comparative in depth RNA sequencing of *P. tricornutum*'s morphotypes reveals specific features of the oval morphotype. *Sci. Rep.* 8:7. doi: 10.1038/s41598-018-32519-7

Prestegard, S., Erga, S., Steinrücken, P., Mjøs, S., Knutsen, G., and Rohloff, J. (2015). Specific Metabolites in a *Phaeodactylum tricornutum* Strain Isolated from Western Norwegian Fjord Water. *Mar. Drugs* 14:9. doi: 10.3390/md14010009

Oliveros, J. C. (2007). *VENNY. An Interactive Tool for Comparing Lists with Venn Diagrams.* Available online at: https://bioinfogp.cnb.csic.es/tools/venny/index.html.

Rastogi, A., Maheswari, U., and Dorrell, R. G. (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. doi: 10.1038/s41598-018-23106-x

Rastogi, A., Vieira, F. R. J., Deton-Cabanillas, A. F., Veluchamy, A., Cantrel, C., Wang, G., 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.* 2, 347–363. doi: 10.1038/s41396-019-0528-3

Remmers, I. M., D'Adamo, S., Martens, D. E., Vos, R. C. H., de, Mumm, R., et al. (2018). Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Res.* 35, 33–49.

Rushforth, S. R., Johansen, J. R., and Sorensen, D. L. (1988). Occurrence of *Phaeodactylum tricornutum* in the Great Salt Lake, Utah, USA. *Great Basin Naturalist* 48, 324–326.

Scarsini, M., Marchand, J., Schoefs, B., Jacob-Lopes, E., Queiroz, M. I., and Zepka, L. Q. (2020). *Carotenoid Overproduction in Microalgae: Biochemical and Genetic EngineeringIn: Pigments from Microalgae Handbook*. Cham: Springer International Publishing, 81–126.

Scarsini, M., Thurotte, A., Veidl, B., Amiard, F., Niepceron, F., Badawi, M., et al. (2021). Metabolite Quantification by Fourier Transform Infrared Spectroscopy in Diatoms: Proof of Concept on Phaeodactylum tricornutum. *Front. Plant Sci.* 12:756421. doi: 10.3389/fpls.2021.756421

Schoefs, B., and Bertrand, M. (2005). in *Handbook of photosynthesis*, Vol. 2, ed. M. Pessarakli (Boca Raton: Taylor & Francis), 37–54.

Song, Z., Lye, G. J., and Parker, B. M. (2020). Morphological and biochemical changes in *Phaeodactylum tricornutum* triggered by culture media: Implications for industrial exploitation. *Algal Res.* 47:101822. doi: 10.1016/j.algal.2020.101822

Stanley, M. S., and Callow, J. A. (2007). Whole cell adhesion strength of morphotypes and isolates of *Phaeodactylum tricornutum* (Bacillariophyceae). *Eur. J. Phycol.* 42, 191–197. doi: 10.1080/09670260701240863

Tesson, B., Gaillard, C., and Martin-Jézéquel, V. (2009). Insights into the polymorphism of the diatom *Phaeodactylum tricornutum* Bohlin. *Bot. Mar.* 52:12. dosi: 10.1515/BOT.2009.012

Vanier, G., Hempel, F., Chan, P., Rodamer, M., Vaudry, D., Maier, U. G., et al. (2015). Biochemical Characterization of Human Anti-Hepatitis B Monoclonal Antibody Produced in the Microalgae *Phaeodactylum tricornutum*. *PLoS One* 10:e0139282. doi: 10.1371/journal.pone.0139282

Vanier, G., Stelter, S., Vanier, J., Hempel, F., Maier, U. G., Lerouge, P., et al. (2018). Alga-Made Anti-Hepatitis B Antibody Binds to Human Fcγ Receptors. *Biotechnol. J.* 13:1700496. doi: 10.1002/biot.201700496

Vartanian, M., Desclés, J., Quinet, M., Douady, S., and Lopez, P. J. (2009). Plasticity and robustness of pattern formation in the model diatom *Phaeodactylum tricornutum*. *New Phytol.* 182, 429–442. doi: 10.1111/j.1469-8137.2009.02769.x

Volcani, B. E. (1981). "Cell wall formation in diatoms: morphogenesis and biochemistry," in *Silicon and Siliceous Structures in Biological Systems*, eds T. Simpson and B. Volcani (New York, NY: Springer), 157–200. doi: 10.1007/978-1-4612-5944-2_7

Willis, A., Chiovitti, A., Dugdale, T. M., and Wetherbee, R. (2013). Characterization of the extracellular matrix of *Phaeodactylum tricornutum* (Bacillariophyceae): structure, composition, and adhesive characteristics. *J. Phycol.* 2013:12103. doi: 10.1111/jpy.12103

Xie, J., Bai, X., Lavoie, M., Lu, H., Fan, X., Pan, X., et al. (2015). Analysis of the Proteome of the Marine Diatom *Phaeodactylum tricornutum* Exposed to Aluminum Providing Insights into Aluminum Toxicity Mechanisms. *Environ. Sci. Technol.* 49, 11182–11190. doi: 10.1021/acs.est.5b03272

Yang, Z.-K., Ma, Y.-H., Zheng, J.-W., Yang, W.-D., Liu, J.-S., and Li, H.-Y. (2014). Proteomics to reveal metabolic network shifts towards lipid accumulation following nitrogen deprivation in the diatom. *J. Appl. Phycol.* 2014:10. doi: 10.1007/s10811-013-0050-3

Young, C. N. J., Sinadinos, A., Lefebvre, A., Chan, P., Arkle, S., Vaudry, D., et al. (2015). A novel mechanism of autophagic cell death in dystrophic muscle regulated by P2RX7 receptor large-pore formation and HSP90. *Autophagy* 11, 113–130. doi: 10.4161/15548627.2014.994402

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chuberre, Chan, Walet-Balieu, Thiébert, Burel, Hardouin, Gügi and Bardor. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership